**PRELIMINARY DRAFT – PLEASE DO NOT CITE WITHOUT PERMISSION**
**COMMENTS WELCOME**

# What Drives Forest Leakage?

January 17, 2016

Kathy Baylis[1], Don Fullerton[1], and Payal Shah[2]

1. University of Illinois at Urbana-Champaign
2. Okinawa Institute of Science and Technology Graduate University

baylis@illinois.edu; dfullert@illinois.edu; payal.shah@oist.jp

**Abstract:** Setting aside forested land for protection can lead to deforestation elsewhere, defined as “leakage”. Under some conditions, it can induce more forest cover near the protected area (“negative leakage”). We develop an analytical general equilibrium model to solve for leakage as a function of key economic variables, and we use it to generate testable hypotheses. We test these hypotheses for Indonesia using landsat data from 2000 and 2012 – before and after eight new national parks were protected in 1999-2004. We estimate deforestation near these parks compared to other unaffected areas, and how it depends on local economic characteristics. To take the spatial nature of deforestation into account, we generate our counterfactual by matching over characteristics of the parcel itself and of its neighbors further away. Conservation agents and social planners can use these results to choose locations of protected areas to help maximize total tree cover.

**Keywords**: Protected areas; deforestation; leakage; matching estimators; general equilibrium; Indonesia; green halo.

* Authors listed alphabetically; lead authorship shared equally. For helpful comments, we thank Katherine Sims, Daniel Karney, and participants at various seminar and conference presentations.

To protect an ecosystem, governments frequently limit deforestation or other resource extraction within a designated conservation zone (Joppa and Pfaff 2010). Globally, 18.4 million square km of marine or terrestrial areas have been set aside primarily for nature and biodiversity conservation (International Union for Conservation of Nature (IUCN) 2010). One criticism is that restricting activity in one area often leads to an increase in ecologically-damaging extraction outside the protected zone, which reduces the effect of the conservation policy.

To study these indirect effects, we borrow wording from other environmental policies. For example, a unilateral emissions policy can achieve direct emission reductions, but that amount can be partially offset by induced changes in emissions elsewhere, called "leakage". Under some conditions, that leakage could be 100% of the emission reduction, implying the policy has no net effect; under other conditions leakage could be negative, so total abatement exceeds direct abatement induced by the policy. In this paper, analogously, we study forest cover and the way that a direct set-aside of protected forest could be partly offset by induced reductions in forest cover elsewhere ("forest leakage"). We show conditions under which that leakage could be negative, with total additional forest cover greater than the initial set-aside, and we test these hypotheses empirically.

While forest leakage is analogous to emissions leakage, the policies and induced indirect behaviors are different. Forest leakage could more than offset any direct conservation gains, if the establishment of a protected area shifts all planned deforestation to other areas and induces anticipatory logging by land users worried about losing their future logging rights (Wunder 2008). Bode *et al.* (2015) use a bio-economic model of forest leakage to illustrate the tradeoffs between human well-being and biodiversity loss owing to possible positive leakage, and they show how such tradeoffs can be minimized if leakage is taken into account during conservation planning. Conversely, establishing protected areas could reduce nearby deforestation or extraction by blocking new roads or by inducing economic growth in non-extractive sectors (Herrera *et al.* 2015). Understanding which characteristics are likely to generate positive or negative leakage could allow policy-makers to site parks that will protect the most total forest.

In this paper, we first develop a general equilibrium model to address what economic factors affect the direction and magnitude of forest leakage. We then test the hypotheses derived from this model and estimate the effect of these factors on leakage in Indonesia. To take the spatial nature of deforestation into account, we compare deforestation in nearby areas against spatially weighted counterfactuals that account for spatial characteristics.

Deforestation in Indonesia is a particular concern because it increases carbon emissions and threatens the survival of numerous endangered species such as orangutans, elephants, and tigers. Indonesia is home to the third largest area of tropical rain forest in the world, but this forest is under threat (United Nations, 2016). Indonesia has the third highest rate of forest loss due to agricultural expansion and unsustainable commercial logging. Between 1990 and 2000, the annual rate of deforestation was 1.61%, and it increased to 1.91% between 2000 and 2005 (Olsen and Bishop, 2009). In response, Indonesia increased the total land area under protection from 10% in 1990 to 14% by 2010. In May 2013, it implemented the world's largest project under the U.N.'s Reducing Emissions from Deforestation and Degradation (REDD+); see Butler (2013). To estimate the effect of these and other potential conservation projects, one must account for the policy's impacts on areas outside the protected zone.

The primary drivers of leakage are likely economic. Since land is a limited resource, a policy that restricts the use of land in one region is likely to induce reallocation of land in other related areas (Schwarze *et al.* 2002). Leakage can also result from the relocation of indigenous communities from protected areas or by inducing preemptive clearing of forest (Armsworth *et al.* 2006; Brokington and Igoe 2006; Oliveira *et al.* 2007; Wittemeyr *et al.* 2008). Gan and McCarl (2007) use a computable general equilibrium model to illustrate the effect of key factors that determine transnational leakage such as the price elasticities of supply and demand for forestry products and the degree of cooperation in forest conservation across countries.

Many impact evaluations report that an increase in protection in one area displaces deforestation activities to other areas (Ferraro 2002; Oliveira *et al.* 2007; Meyfroidt and Lambin 2009). Conversely, several studies have found negative forest leakage (or a "green halo" effect), where protection increases the forest conservation on adjacent lands (Honey-Rosés *et al.* 2011; Pfaff *et al.* 2014; Gaveau *et al.* 2009). Thus, empirical estimates of leakage are mixed.

Previous studies that model leakage account for the effects on forest cover through changes in prices as well as demand and supply conditions in single markets. Wear and Murray (2004) find that a mandated reduction in logging activity on public lands in the U.S. Pacific Northwest resulted in intensified timber harvesting on nearby private lands. Wu (2000) develops a conceptual framework to estimate potential leakage within the Conservation Reserve Program (CRP); he finds that about 20% of the acreage enrolled in the CRP was offset by expansion of cropland cultivated elsewhere. These studies do not account for the effect that leakage has on prices and production decisions in other sectors, including both extractive and non-extractive sectors. In this paper, we investigate economic variables that play key roles in determining the direction and intensity of leakage.

We develop a simple general equilibrium model with two inputs (land and labor) and four outputs (eco-services, timber, agriculture, and manufacturing). We use it to ask how local characteristics of markets affect the extent and type of leakage, and we derive eight theorems. We then calibrate parameters, so that each theorem yields a particular hypothesis for Indonesia. For our first theorem, we derive the conditions under which leakage is positive or negative. For our other theorems, we show how leakage depends on each key parameter (all else equal). For example, leakage will be larger in cases where the land taken for protection comes relatively more from what would have been used for agriculture rather than from timber production. We also show leakage will be larger when the fraction of other nearby land initially in timber is relatively high (or equivalently, the fraction initially used in agriculture is low). Leakage is larger when nearby agricultural products have relatively inelastic demand, and it is smaller when nearby forest products have relatively inelastic demand. We then test those hypotheses.

To be able to solve this general equilibrium model analytically, we limit the number of sectors, factor inputs, and other complications such as heterogeneity. The theoretical model has one agricultural output, one tree product, one type of land, and one type of labor, along with assumptions of perfect competition and constant returns to scale. Other considerations may affect the amount of forest leakage, but we abstract from them to focus on effects of economic drivers that we believe to be most important, including demand elasticities, factor shares, and other key differences across parks. We do capture heterogeneity across districts of Indonesia, however, since our model can be taken to represent a locality that faces particular demand elasticities for its primary agricultural and tree products that are produced using particular factor shares, so that we predict the forest leakage in that locality. We then test the model using data across Indonesia for different localities that produce different goods using different factors near different parks. We thus test only some of the drivers of forest leakage, while controlling for other drivers of leakage by including other control variables in the regressions.

Indonesia was successful in reducing deforestation within its new protected areas. In earlier work, Shah and Baylis (2015) found that the new protected areas had 1.1% more forest cover remaining than did their control counterparts. When we consider leakage here, we find that the new protected areas *increase* total forest cover in nearby areas by 1% (relative to controls). Thus, on average, the establishment of protected areas decrease land conversion on nearby parcels (negative leakage). That said, these national numbers mask a great deal of heterogeneity. Effectiveness of new protected areas established after 1999 range from –3.4% to +5.3%, and estimates of leakage range from –10.3% to +7.5%.

Here, we explain these differences in leakage using observed variation in local economic characteristics of the forest and agricultural sectors. For example, districts contiguous to Tesso Nilo National Park in Sumatra produce a relatively large fraction of agricultural commodities that have inelastic demand. These economic conditions in our first theorem imply that an increase in land protection leads to an increase in deforestation activities in nearby districts. Indeed, we find statistically significant evidence of positive leakage surrounding that park. Conversely, districts contiguous to Sebangau National Park in Kalimantan produce a relatively smaller proportion of such goods, so their output faces more elastic demand. As a result, our theory and significant evidence both suggest a *decrease* in deforestation pressures in areas near this new protected area. We test these relations for the regions surrounding all of the newly established parks. However, since we compare these affected areas near each new park to other controls that are further away within Indonesia, we measure only this crucial nearby portion of leakage, not the effects of these new Indonesian parks on deforestation worldwide.

Our empirical results generally corroborate our hypotheses about effects of key economic indicators on leakage. Given differing parameters around each park, our first theorem/hypothesis correctly predicts one of the two areas with positive leakage, and five of the six with negative leakage. Also, as predicted by other theorems, we find lower leakage near parks when more of that protected area comes from forest land than from agriculture. Also as predicted, we find less leakage if nearby districts produce agricultural products with a higher price elasticity of demand or produce timber products that have lower demand price elasticity. Finally, we find relatively higher deforestation pressures in nearby districts that have larger share of labor relative to land.

This paper makes several contributions to the existing literature on impact evaluation of protected area policies. First, using a general equilibrium framework, we generate predictions of how economic variables affect leakage – reductions in nearby forest cover. While past studies have estimated leakage for one particular region or one particular sector of the economy, they have not estimated how different economic settings generate different signs and magnitudes of leakage (with the exception of recent work by Pfaff *et al.* 2014). Our study also incorporates spatial considerations that capture impacts on deforestation based on characteristics of neighboring areas. Particularly in the context of future REDD projects, conservation planners may be able to use these results to target the placement of future conservation sites to maximize their conservation effect both inside and outside the protected area.

## 1. A Simple General Equilibrium Model

In a developing country with finite land and forest resources, any policy to set aside land for conservation must take land away from agriculture and forest-based production activities. Shrinking those sectors may directly affect those outputs by removing land, reallocating labor, or both. The policy therefore affects relative prices, which affects production decisions throughout the economy. We construct an empirically testable general equilibrium model that captures these essential elements and remains simple enough to solve analytically. In this section, we develop a basic model that represents an equilibrium in all input and output markets. We differentiate these equations and then solve simultaneously for solutions that show the effect on all prices and quantities from a small policy shock such as placing additional land into forest reserves.

Consider an open economy, not necessarily a small open economy, with constant returns to scale production of four outputs using two inputs. One input is land and forest resources, $R$, and the other is called labor, $L$. This second input represents all inputs other than land, so it could be interpreted as labor, human capital, physical capital, or a composite of all such inputs.

The first output is “environment and ecosystem services”, $E$, produced by government using land and forest resources set aside for conservation (in amount $R_E$). We treat $E$ as a public good in utility, rather than as a private good for purchase. Thus, households do not choose the amount of $E$, but they may benefit from visiting the park or from non-use existence value. Moreover, the travel industry could benefit from $E$ for ecotourism; drug companies could use it for biodiversity to develop new pharmaceuticals; non-profit organizations may value conservation easements; and $E$ might be used for sequestration credits.

The other three goods are produced and sold in competitive markets. One is a sustainable forest-based product such as “timber”, $T$, produced using land and forest resources (in amount $R_T$) and labor (in the amount $L_T$). Good $T$ includes sustainably harvested second-growth timber as well as non-timber forest products. The private demand for $T$ arises outside the economy from those who buy sustainable timber products or sequestration credits from the sustainable use of forests. Production of goods $E$ and $T$ both retain tree cover (in the amount $R_E + R_T$).

The third good our economy produces is an agricultural good, $A$, which requires the clearing of forest, the potential harvest of old-growth timber, mining products, palm oil, and other agricultural exports. We suppose that production of this good also uses land resources, in the amount $R_A$, and the other composite input called labor, in amount $L_A$, earning wage $w$. The fourth and final output is "manufacturing", $M$, produced using only labor, $L_M$, earning wage $w_M$. This good represents all other economic output, so it effectively includes food and services as well as manufactured goods. The production functions and zero profit conditions are:[1]

$$E = R_E$$

$$T = T(R_T, L_T) \qquad TP_T = R_T P_R + L_T w$$

$$A = A(R_A, L_A) \qquad AP_A = R_A P_R + L_A w$$

$$M = L_M \qquad MP_M = L_M w_M$$

where $P_i$ is the market price of good $i$ (for $i = T, A, M,$ and $R$). The elasticity of substitution in production of timber is $\sigma_T$, and the one in agriculture is $\sigma_A$. Each of the private industries must break even, with no excess profits, so the first zero profit condition says that revenue from sale of timber ($TP_T$) must match the cost of producing it ($R_T P_R + L_T w$). Similarly, for agriculture, revenue $AP_A$ must match costs $R_A P_R + L_A w$. With only one input in $M$, zero profits implies that the wage $w_M$ must match the price of output $P_M$.

We assume that labor is perfectly mobile between agriculture and timber production, presuming that both are in rural areas. This rural labor, $L_{TA} \equiv L_T + L_A$, earns a single wage $w$. Labor is not perfectly mobile between this rural area and the urban area, where labor used to produce $M$ receives wage $w_M$. To represent imperfect mobility of labor, we use the function $L = L(L_{TA}, L_M)$, where $\sigma_L$ is the elasticity of $L_{TA}/L_M$ with respect to $w_M/w$, and where $LP_L = L_{TA}w + L_M w_M$ is the corresponding zero profit condition.[2]

[1] A general production function with one input could be written as $E = E(R_E)$ or $M = M(L_M)$, but constant returns to scale production means that each function must be linear. Moreover, we can define a unit of output as the amount produced using one unit of the input, which yields the simpler forms of production shown.

[2] The function $L = L(L_{TA}, L_M)$ looks like a production function, but we do not need to solve for "output" $L$ or its "price" $P_L$. Instead, this function for $L$ is just a convenient way to moderate movement between $L_{TA}$ and $L_M$.

Then the two resource constraints are:

$$\bar{R} = R_E + R_T + R_A \quad \text{and} \quad \bar{L} = L_{TA} + L_M \ .$$

For simplicity, assume that $T$ and $A$ are produced only for export, which means that this economy can import some of the manufactured good, in amount $M^I$. Consumption is then the sum of domestic production and imports: $M^C = M + M^I$. Balance of payments requires that the value of exports equals the value of imports:

$$TP_T + AP_A = M^I P_M$$

The government can remove land from private production of $T$ or $A$ and allocate it to production of $E$, the public good.[3] This $E$ could provide international as well as domestic benefits. Domestic (local) utility is a homothetic function of manufactured goods and the public good, $U = U(M^C; E)$. Given $E$, many identical local households choose $M^C$ to maximize utility subject to their budget, $M^C P_M = (\bar{R} - R_E) P_R + L_{TA} w + L_M w_M$. Since all functions are homothetic of degree one, we have no need to specify the number of firms or the number of households. That is, we can define all of these inputs and outputs as amounts per household. We assume that this local economy's goods $T$ and $A$ are not the same as others' timber or agricultural products. Thus, this economy faces downward sloping demands for its exports. We have no need to specify the full utility-maximizing behavior of the rest of the world (ROW). Rather, we merely suppose that $\eta < 0$ is their price elasticity of demand for good $A$, and $\varepsilon < 0$ is their price elasticity of demand for good $T$ from this economy.[4]

Good $M$ is undifferentiated and traded worldwide, and so domestic producers face a fixed world price $P_M$. Then, because production is linear ($M = L_M$), the zero profit condition effectively fixes the urban wage. We use that price as numeraire, so $w_M = P_M$ never changes.

Next, suppose a policy shock brings additional land under protection. That is, the policy increases $R_E$, the amount of land for conservation and eco-services, $E$, which results in less land

---

[3] In our simple model, we have no need to identify a separate government budget. Essentially, the one budget constraint represents a consolidation of consumer and government budgets. The government may use lump sum taxes to purchase more land for production of $E$. It may even sell some $E$ and return the proceeds to consumers (or charge a smaller lump sum tax). Alternatively, the government could just confiscate the land for production of $E$. Thus, we do not need to consider a price for eco-services, $P_E$ .

[4] Later, we differentiate all equations in the model and define a variable with a hat as the proportional change in that variable (for example, $\hat{T} \equiv dT/T$). Then our elasticity specification here just means that the change in export demand for $T$ is $\hat{T} \equiv \varepsilon \widehat{P_T}$, and the change in export demand for $A$ is $\hat{A} \equiv \eta \widehat{P_A}$ .

available for production of $T$ or $A$. Thus, we expect the input price of land to rise relative to the price of labor. Since both $T$ and $A$ use the land resource $R$, those output prices also must rise for firms to break even (and the output that is more land-intensive will need its price to rise more). The change in each such output is moderated by a downward-sloping demand. If the ROW demand were inelastic, then the economy may receive an increase in export revenue and be able to import more $M$. In that case, households could achieve greater utility, as the removal of land from production effectively allows the economy to exploit its own market power over the goods it exports. If those ROW demands are elastic, however, then this economy's higher cost of producing exports means less export revenue and less import of $M$.

In this static model, the best interpretation is not that farmers or firms react to this policy shock by converting agricultural land to timber, or vice versa. Rather, the comparison is counterfactual. We look years after the policy shock to see whether lands that would have been converted to agriculture are instead kept forested, or vice versa.

### *1.1 Linearization and Solution*

We differentiate all these equations to linearize the model, where Appendix 1 identifies the resulting 19 linear equations and matching number of unknowns (new prices and quantities). Then, we solve the system of 19 linear equations for the effects of a small exogenous increase in the resource set aside for eco-services ($\hat{R}_E \equiv \frac{dR_E}{R_E} > 0$). The closed-form solution for the change in the equilibrium amount of the resource used in production of timber, $\hat{R}_T$, is:

$$\hat{R}_T = -\frac{\lambda_{RE}N}{\lambda_{RT}N+\lambda_{RA}D}\hat{R}_E. \tag{20}$$

where

$$N = -(\varepsilon\theta_{TR} - \sigma_T + \theta_{TR}\sigma_T)(\eta\lambda_{LA} + \lambda_{LM}\sigma_L) + \varepsilon\lambda_{LT}\sigma_T + \varepsilon\lambda_{LA}\theta_{AR}(\eta + \sigma_A) \tag{21}$$

$$D = (\varepsilon + \sigma_T)\theta_{TR}\eta\lambda_{LT} + \eta\lambda_{LA}\sigma_A - (\eta\theta_{AR} - \sigma_A + \theta_{AR}\sigma_A)(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT}) \tag{22}$$

and where $\lambda_{ij}$ is the fraction of factor $i$ ($i$=$R$,$L$) initially in sector $j$ ($j$=$E$,$T$,$A$,$M$). Thus, we have $\lambda_{RA} + \lambda_{RT} + \lambda_{RE} = 1$ and $\lambda_{LA} + \lambda_{LM} = 1$. Also, $\theta_{ji}$ denotes the factor share in sector $j$ ($j$=$T$,$A$) for each factor $i$ ($i$=$R$,$L$), so $\theta_{AR} + \theta_{AL} = 1$ and $\theta_{TR} + \theta_{TL} = 1$. The signs of equations (21) and (22) are indeterminate, and so the sign of equation (20) is also indeterminate.

The policy shock sets aside some land for eco-services ($dR_E > 0$) and retains it in natural vegetation including tree cover. The government takes a fraction $\alpha$ of that land $dR_E$ from what would become agricultural, and a fraction $(1–\alpha)$ from what would have remained in sustainable tree cover. Thus, the direct addition to tree-covered natural area is simply $\alpha dR_E$. This term captures the direct "additionality" of the policy to create the protected area. Then leakage reflects private market re-allocations between remaining treed land and agricultural land.

Figure 1 may clarify. The first and primary effect of this addition to national park, $dR_E$, is the direct additional forest, $\alpha dR_E$, the portion of the land added to the national park that is taken from what would otherwise have become agricultural (with a red rectangle around it in Figure 1). "Leakage" pertains to the remaining privately-owned land. Positive leakage means that remaining ag land expands: if the black vertical line shifts left, then the net addition to forest cover is less than $\alpha dR_E$. Conversely, negative leakage means that remaining private treed land expands. It is already reduced by $(1–\alpha)dR_E$, but if it re-expands to encroach on ag land, then the overall addition to forest cover is more than the direct set-aside, $\alpha dR_E$.

In the emissions leakage literature, the absolute amount of leakage outside the policy zone is normally expressed as a fraction of the amount of protection inside the policy zone. For analogous definitions here, we take the absolute amount of remaining private land that switches from $T$ to $A$, and divide by $\alpha dR_E$ , the government's direct addition to tree cover. For the numerator of that ratio, we define the quantity of leakage as $– [(1–\alpha)dR_E + dR_T]$. The first term is the preserved land taken directly from private tree cover. If the overall equilibrium reduction in private tree cover ($dR_T < 0$) exactly matches the land taken directly from tree cover, $(1–\alpha)dR_E$, then leakage is zero. To the extent that the size of the equilibrium change ($dR_T < 0$) exceeds the amount preserved from tree cover, $(1–\alpha)dR_E > 0$, then leakage is positive. Conversely, if the equilibrium reduction in private tree cover is smaller than the amount taken for preservation from tree cover, then leakage is negative.

Then the leakage fraction, $\psi$, is the ratio of that leakage over the direct additional tree cover in the natural area set aside by the policy:

$$\psi \equiv -\frac{(1-\alpha)dR_E + dR_T}{\alpha dR_E}$$

With rearrangement and conversion to our hat notation, we can write this leakage ratio as:

$$\psi = -\left[\frac{(1-\alpha)}{\alpha} + \frac{\widehat{R_T}}{\alpha \widehat{R_E}} \frac{R_T}{R_E}\right] = -\left[\frac{(1-\alpha)}{\alpha} + \frac{\widehat{R_T}}{\alpha \widehat{R_E}} \frac{\lambda_{RT}}{\lambda_{RE}}\right] = -\left[\frac{(1-\alpha)}{\alpha} - \frac{\lambda_{RT} N}{\alpha(\lambda_{RT} N + \lambda_{RA} D)}\right] \quad (23)$$

where the last expression follows from the solution for $\hat{R}_T$ from equation (20) and rearranging.

The first term is the fraction of the land taken from tree cover (1–$\alpha$) over the fraction taken from agriculture ($\alpha$). From here on, we call this term the "set-aside effect" (SAE). It is a negative effect on leakage: more land taken from private tree cover (1–$\alpha$) means that remaining private lands would have to switch more from agriculture to replace that timber production. Since the sign on $\hat{R}_T$ is indeterminate, the sign on the second term is also indeterminate. However, net leakage can be negative if the magnitude of the negative SAE is large enough.

### *1.2 Intuitive Special Cases*

A couple of special cases might help with intuition. First, one may ask, when is leakage exactly 100%, such that any deforestation prevented in one location becomes deforestation elsewhere instead? In equation (23), the leakage rate $\psi$ = 1 whenever $D$ = 0. And from (22), we can show that $\eta = \sigma_A = 0$ is sufficient for $D$ = 0. If market demand is unchanged for agricultural output ($\eta$=0), and if that production requires unchanged use of land ($\sigma_A$=0), then any agricultural land taken for the park must be replaced by use of other private land. In other words, those who expect 100% leakage might not be thinking about behavioral reactions to the raised price of land used in agriculture.

Second, when is leakage exactly zero? From (23), $\psi$ = 0 whenever $(1-\alpha) = \frac{\lambda_{RT} N}{(\lambda_{RT} N + \lambda_{RA} D)}$, which might occur by coincidence for particular values of parameters. Yet that condition for zero leakage is certainly satisfied when $\alpha$=1 and $\varepsilon = \sigma_T$=0 (so that $N$=0). That is, leakage is zero when all land for the new park is taken from agriculture ($\alpha$=1), and markets demand unchanged land for tree products ($\varepsilon=\sigma_T$=0), because then no land gets switched back to agriculture.

### *1.3 Calibration of Parameters*

The next subsection below uses the solution of our theoretical model in equation (23) to derive eight theorems that show how leakage depends on parameters. For each theorem, we then insert parameter values in order to state a specific hypothesis for Indonesia. To do so, we first explain here our derivation or calibration of those parameters.

Table 1 shows the parameters we use in illustrations for the whole country. Most of these parameters are calculated from our data described below for the thousands of parcels used in our regressions. For most of the parameters in Table 1, we simply average the corresponding variable across all of Indonesia. For example, we use the "1990 Indonesian Spatial Plan" to identify whether each 3 km by 3 km parcel is allocated to be used for conservation, for sustainable tree products, or for agriculture (to calculate the average $\lambda_{RE}$, $\lambda_{RT}$, and $\lambda_{RA}$, in the first three rows). The underlying data for most other parameters in Table 1 are described in section 2.3 below.

Data are not available for the elasticities of substitution in the last three rows, however, so values must be assumed. For production, we note that the flexibility to change factor input ratios must depend on the time frame allowed for such substitution. In our case, most of the new parks were established in 2004, and our data show forest cover only eight years later (in 2012). Therefore, we assume relatively low elasticities of substitution in production ($\sigma_A = \sigma_T = 0.2$). In order to assume somewhat limited mobility of labor, we use 0.3 for $\sigma_T$ (the elasticity of $L_{TA}/L_M$ with respect to $w_M/w$). Later, we discuss the sensitivity of results to these assumptions.

### *1.4 Testable Hypotheses*

Using this model and the solutions above, we prove eight theorems. We then use data from Indonesia to test the applicability of the model. For the first theorem, we derive conditions under which leakage is positive or negative. For the other theorems, we differentiate (23) with respect to key parameters to predict the effect of each parameter on leakage. We use these theorems to generate testable hypotheses, which we test statistically. That is, we investigate the effect on leakage of empirical variation across areas of Indonesia in variables such as the share of land taken from agriculture, key factor shares in production, and each of the demand elasticities.

*Theorem 1: Leakage is negative if and only if* $\alpha < \frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)}$ . Proof: Appendix 2.

*Hypothesis 1: Leakage is negative near the new Indonesian parks where* $\alpha < \frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)}$ .

If the size of the equilibrium change in private tree cover, $dR_T < 0$, is smaller than the amount taken away for preservation, $(1–\alpha)dR_E > 0$, then leakage is negative.

To test this first hypothesis below, we undertake a new calibration for just the districts contiguous to each new national park. We use this park-specific calibration to calculate the predicted sign of leakage near each park (from the inequality in the first hypothesis). Then, finally, we see if measured leakage around each park has the predicted sign.

To derive the next seven theorems, we differentiate the solution for leakage in (23) with respect to $\alpha$, $\eta$, $\varepsilon$, $\lambda_{RA}$, $\lambda_{RT}$, $\theta_{AL}$ and $\sigma_L$.[5] First, consider $\alpha$, the share of land taken from what would have become agriculture. Since $\alpha$ enters both terms in (23), a higher $\alpha$ has two opposite effects. The first term (the SAE) is a negative term in leakage, $-(1-\alpha)/\alpha$, made smaller when $\alpha$ is larger (which has a positive effect on leakage). When taking more land directly from agriculture, $\alpha$, the private market tends to reallocate more land from timber back into agriculture. The second term in (23) is positive, also made smaller when $\alpha$ is larger (a negative effect on leakage). The next theorem shows when the first effect dominates.

*Theorem 2: Leakage is larger in cases with a higher fraction of set aside taken directly from agriculture* ($\alpha$), *if and only if* $\frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)} < 0$. Proof: Appendix 2.

*Hypothesis 2: Using our data averaged across Indonesia, we find* $\frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)} < 0$, *so we expect leakage is larger near new parks that take more land from agriculture.*

Figure 2 uses our parameterized values for all of Indonesia in Table 1 to illustrate numerically how a larger $\alpha$ implies larger leakage (holding constant the other parameters). Intuitively, a larger $\alpha$ indicates that more land for preservation is taken directly away from agriculture, so the private market is predicted to reallocate more land from what would have remained timber back into agriculture (a positive effect on leakage).

*Theorem 3: Leakage is smaller in cases with a more elastic demand for agricultural output* (*larger* $|\eta|$ ) *if and only if* $D\big(\varepsilon\lambda_{LA}\theta_{AR} - \lambda_{LA}(\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)\big) - N\big((\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)\lambda_{LT} + \lambda_{LT}\sigma_T + \lambda_{LA}\sigma_A - \theta_{AR}(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT})\big) < 0$. Proof: Appendix 2.

*Hypothesis 3: This inequality holds for data averaged across Indonesia, so we expect leakage is smaller in areas that face a more elastic demand for their agricultural output.*

The price elasticity of demand for the agricultural good is $\eta$, negative by definition. We find that the sign of the derivative $\partial\psi/\partial\eta$ is positive based on our data for Indonesia. Hence, we expect to find that leakage is larger in areas where $|\eta|$ is smaller (more inelastic demand for *A*). Intuitively, when land is set aside for *E*, remaining available land becomes scarce and its price rises. This effect raises the price of both goods $T$ and $A$ (relative to the numeraire wage

[5] We use proxies for data on these parameters to test hypotheses below. We do not derive theorems for the effect of other parameters in (23), because good proxies are not available for testing them (i.e. $\theta_{TR}$, $\theta_{TL}$, $\theta_{AR}$, $\sigma_A$, and $\sigma_T$).

or price of $M$). All else equal, a less elastic demand for $A$ then means that the market shifts more remaining private land from $T$ to $A$ to meet demand (a positive effect on leakage).

Conversely, as stated in Hypothesis 3 and shown for our parameters in Figure 3, a larger demand elasticity $|\eta|$ means that the higher price of $A$ leads to shrinking production of $A$, and private land can shift from agriculture to forest production (a negative effect on leakage).

*Theorem 4: Leakage is smaller in cases with a less elastic demand for sustainable tree products (smaller $|\varepsilon|$ ) if and only if* $D\big(-\theta_{TR}(\eta\lambda_{LA}+\lambda_{LM}\sigma_L)+\lambda_{LT}\sigma_T+\lambda_{LA}\theta_{AR}(\eta+\sigma_A)\big)-N\big(\theta_{TR}\eta\lambda_{LT}-\lambda_{LT}(\eta\theta_{AR}-\theta_{AL}\sigma_A)\big)<0$. Proof: Appendix 2.

*Hypothesis 4: This inequality holds for data averaged across Indonesia, so we expect leakage is smaller in areas facing less elastic demand for their tree products.*

The price elasticity of demand for timber is $\varepsilon$, negative by definition. Since we find that the derivative $\partial\psi/\partial\varepsilon$ is negative based on our data averaged over all of Indonesia, as seen in Figure 4, we expect that leakage near new parks is smaller in districts that produce tree products having relatively inelastic demand. Intuitively, when the price of $T$ rises, inelastic demand leads to a reallocation of private resource $R$ from agriculture to timber to meet this inelastic demand. Hence leakage is smaller or more negative.

*Theorem 5: Leakage is smaller in cases with a larger initial allocation of land resource $R$ to agricultural production, $\lambda_{RA}$, if and only if* $ND>0$. Proof: Appendix 2.

*Hypothesis 5: This inequality holds for data averaged across Indonesia, so we expect leakage is smaller in areas with a larger initial land in agriculture.*

We find that $\partial\psi/\partial\lambda_{RA}$ is negative based on our data for Indonesia, which indicates that a larger $\lambda_{RA}$ implies smaller leakage (as seen in Figure 5). If a large amount of land resource $R$ is initially allocated to agricultural production, then following the policy shock, more ag land is *available* to switch from production of $A$ to $T$; hence, leakage is smaller (and possibly negative).

*Theorem 6: Leakage is larger in cases with a larger initial allocation of land to timber, $\lambda_{RT}$, if and only if* $ND>0$. Proof: Appendix 2.

*Hypothesis 6: This inequality holds for data averaged across Indonesia, so we expect leakage is larger in areas with a larger initial land in timber or tree products.*

The derivative $\partial\psi/\partial\lambda_{RT}$ is positive based on our data for Indonesia, so a larger $\lambda_{RT}$ implies larger leakage (as seen in Figure 6). If a relatively large amount of land resource $R$ is initially allocated to timber production, then following the policy shock, more of that land is available to switch from production of $T$ to production $A$; hence, leakage is larger.[6]

*Theorem 7: Leakage is smaller in cases with a larger factor share of labor in agriculture,* $\theta_{AL}$*, if and only if* $(\eta + \sigma_A)(D\varepsilon\lambda_{LA} + N(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT})) > 0$. Proof: Appendix 2.

*Hypothesis 7: This inequality holds for data averaged across Indonesia, so we expect leakage is smaller in cases with a larger* $\theta_{AL}$*.*

Figure 7 uses our parameterized values for Indonesia to show how a larger $\theta_{AL}$ implies smaller leakage. Intuitively, a larger $\theta_{AL}$ indicates that agriculture is relatively labor-intensive. Then taking land from agriculture means a decreased need for reallocation of remaining private land from timber to agriculture, which decreases leakage.

*Theorem 8: Leakage is smaller in cases with a larger* $\sigma_L$ *(elasticity of* $L_{TA}/L_M$ *with respect to* $w_M/w$*) if and only if* $N(\eta\theta_{AR} - \theta_{AL}\sigma_A) - D(\varepsilon\theta_{TR} - \theta_{TL}\sigma_T) < 0$. Proof: Appendix 2.

*Hypothesis 8: This inequality holds for data averaged across Indonesia, so we expect leakage is smaller in cases with more mobility.*

Taking land from production of either timber or agriculture raises the available labor relative to land (as seen in Figure 8). When that labor can move more easily to manufacturing, then the wage does not have to fall enough for that labor to be employed locally (mostly likely in the labor-intensive agriculture sector). In that case agriculture does not need to re-expand to re-employ that labor, which implies less leakage.

## 2. Empirical Methods

By 2000, the government of Indonesia had already designated 320,000 sq. km in protected areas, and it designated an additional 15,300 sq. km in the form of 20 new protected areas between 2000 and 2012. Of those, 12 are marine protected areas, and eight are terrestrial national parks. The eight new terrestrial national parks comprise about 10,000 sq. km. and are located in or contiguous to 35 districts; we treat each district as a small, separate economy

[6] Theorems 5 and 6 are obviously related to each other, but they do not make exactly the same point. Because $\lambda_{RT} + \lambda_{RA} + \lambda_{RE} = 1$, a larger $\lambda_{RT}$ does not necessarily mean a smaller $\lambda_{RA}$.

(Burgess *et al.* 2011). We treat these eight recently established protected areas as our measure of the policy shock.

### *2.1 OLS and IV Regressions*

We use a spatial lag model with data from Indonesia to estimate leakage from protected areas established between 1999 and 2004, and we empirically test the hypotheses derived above. Our basic approach is to use a difference-in-difference methodology wherein our dependent variable, deforestation, represents the difference in forest cover between 2000 and 2012. We compare deforestation in the "treatment" parcels that are near these protected areas to controls that are further away (that do not have a protected area "nearby"). Each treatment parcel is matched to a control parcel based on characteristics that affect protection and deforestation decisions (as in Honey-Rosés *et al.* 2011).

We relate this measure of leakage to the district-level economic conditions that reflect demand for goods from agricultural products and timber products ($A$ and $T$), as well as other variables. We use the set of matched parcels to generate a dataset on which we can estimate the signs and significance of the effects of key economic variables on deforestation and leakage. We first estimate seven separate OLS regressions, one for each proxy that we use to represent the seven key economic variables in the theoretical model discussed in section 1,

$$Deforestation = \alpha + \beta(Treatment\ Dummy) + \gamma(Covariates) +$$
$$\delta(Economic\ Variable) + \pi(Treatment\ Dummy \times Economic\ Variable).$$

Here, we regress deforestation per parcel on the proxy for each economic variable interacted with a dummy identifying whether the parcel is a treatment parcel (defined as a parcel within a district contiguous to the park). We control for biophysical characteristics of each parcel such as slope, elevation, forest cover in 2000, and other determinants of deforestation (distance to city and distance to road). We include province dummies to control for the average difference in deforestation across provinces. We test the hypothesis from the theoretical model by looking at the sign of the coefficient on the interaction term. We then estimate the OLS regression with all seven proxy economics variables together. For this regression, we regress deforestation per parcel on all proxy economic variables, their interactions with the treatment dummy, covariate terms, and province dummies.

Previous studies find evidence of spatial dependence in deforestation decisions and land use models (Mertens and Lambin 2000; Anselin 2002; Alix-Garcia 2007; Alix-Garcia *et al.* 2010). These studies show that forest loss has distinct spatial patterns where deforestation in one region is likely to be affected by the deforestation decisions in neighboring regions. However, measuring neighbors' interactions is a difficult task because of possible feedback effects that allow individual observations and observations on neighboring parcels to affect each other simultaneously (Manski 1993; Brock and Durlauf 2001; Moffitt 2001). Robalino and Pfaff (2012) apply the instrumental variable (IV) approach wherein they use exogenously varying topological and ecological characteristics of neighbors' parcels as instruments for the possible neighbors' interactions in deforestation decisions.

Therefore, in a separate IV specification, we also control for a spatial lag of deforestation, using the procedure of Kapoor *et al.* (2007). Here, we instrument for spatial lag of the dependent variable with spatially lagged values of covariates that we include in the matching process.

### *2.2 Identification Strategy*

We focus on 3 km by 3 km parcels. We eliminate parcels with zero forest cover in 2000 and parcels that were designated as protected before 1999. Of the remaining 173,956 parcels, 3057 parcels belong to the eight protected areas newly created between 1999 and 2004.[7] Next, we define the 'nearby' areas that may experience leakage as the districts that are contiguous to the parks. We expect that these areas will experience the greatest impact from the establishment of the protected area. We compare these treated parcels to similar parcels further away but still within the same province.

Because of unobserved spatial heterogeneity and limited factor mobility, we anticipate that effects of land conservation will be largest in districts nearby the removal of land from production. One drawback of this approach is that similar areas further away may still be affected by some of the changes in price induced by the change in protection. With respect to the eight theorems or hypotheses, we do expect price elasticities of demand for agricultural and mining outputs (good $A$) and for sustainable tree products (good $T$) to have effects on leakage primarily nearby but potentially somewhat in more distant areas as well. Limited mobility of labor may imply that the initial factor share of labor in agriculture ($\theta_{AL}$) has effects on leakage that are relatively stronger nearby. Other variables are also more likely to have stronger effects

[7] Of our eight new parks, seven were established in 2004. Only Kerinci Seblat National Park was established in 1999. We include all eight parks because this 1999 park would have no effect on our "initial" forest cover in 2000.

nearby. In any event, what we measure is not the total amount of leakage, but the difference between leakage nearby minus any leakage that occurs further away (in control parcels).

Previous studies use simple distance measures to place zones around protected areas to delineate where they expect to find leakage (Andam *et al.* 2008; Pfeifer *et al.* 2012). In Appendix 3, we present results using such distance measures, where we define "nearby" areas as those that are within a 35 km radius of the new protected areas.

One challenge facing empirical estimation of leakage is that protected areas are not randomly placed across the landscape. Recent empirical studies use methods that formally develop a counterfactual control group to overcome the potential selection bias arising from non-random assignment of protected areas (Ferraro and Pattanayak 2006). Some previous studies use propensity score matching to compare the deforestation rates in protected areas with matched counterfactuals (Joppa and Pfaff, 2010; Gaveau *et al.*, 2009; Andam *et al.*, 2008). In our case with spatial auto-correlation, propensity score matching is inappropriate, because its use of probit estimation would be inconsistent. Instead, to develop a control group, we use parcels that are outside the areas contiguous to the new park but that share similar characteristics with the parcels near the park. Specifically, we select control parcels using the nearest neighbor matching technique of Abadie and Imbens (2006). To control for time-invariant parcel fixed effects, we consider only the change in forest cover from 2000 to 2012. We define as leakage the change in forest cover of parcels in these nearby treatment areas relative to matched controls further away. We then use the matched sample to regress forest loss on economic characteristics that our model suggests would predict the sign and magnitude of leakage.

In generating our counterfactuals, we match over covariates that are likely to affect the selection of a parcel into protection and that affect the extent of forest degradation. Indonesian policymakers target land parcels for protection based on important determinants of deforestation such as slope, elevation, soil type, and peat depth (Wich *et al.* 2011). Other determinants of the probability of deforestation include plot-level accessibility characteristics such as distance to roads, distance to rivers, distance to nearest city, and land use opportunities (Angelsen and Kaimowitz 1999; Pagiola 2000; Deininger and Minten 2002; Alix-Garcia 2007). Further, we expect parcels within the same administrative region and ecological region to be affected by similar deforestation pressures; thus we restrict each match used as a control parcel to be within the same province as the treatment parcel and in the same ecoregion (as defined by the World

Wildlife Fund). To ensure that the type of forest production activity does not vary greatly between treatment and control parcels, we also restrict matches to be in the same land management category. For the areas near the national parks established in 2004, we also include the difference in forest cover between 2000 and 2002 as a covariate to control for the trend in forest cover change prior to the establishment of the new national park.[8]

To assess the quality of the matches, we calculate the normalized difference for each covariate between treated parcels and matched control parcels. We also calculate the normalized differences for biophysical characteristics that are not used in the matching process, to evaluate whether the matching process is successful in identifying physically similar parcels. We match three control parcels for each treated parcel, and we use the bias adjustment procedure to control for match fit (Abadie and Imbens 2006). To test the robustness of final results to the choice of using three matches, we also generate sets of counterfactuals using two and four matches.

Deforestation in any one area is likely to be affected by biophysical characteristics of its neighbors. Results from the Moran's I test of spatial autocorrelation indicate a significant positive correlation of 0.59 (p-stat of 0.00) between the extent of forest cover on a given parcel of land and on its neighboring parcels. We derive spatially weighted values for the above-mentioned characteristics using a queen's contiguity based weights matrix in GeoDa.[9] The belief that deforestation is likely to spread continuously over space justifies defining neighboring parcels by contiguity. To identify control parcels for each treated parcel, we match parcels over their own characteristics and their spatially weighted covariates (Honey-Rosés *et al.* 2011).

Our measure of leakage may be biased if our treatment area includes 3x3km parcels that are located right on the boundary of a newly established park (Honey-Rosés *et al.* 2011). If these parcels are affected by negative or positive spatial spillovers simply from being adjacent to a protected area, then inclusion of these parcels in the treatment group may over- or under-state the impact of protection on leakage. In our analyses, to correct for this potential bias, we remove 3x3 parcels that are immediately adjacent to the boundary of the new park.

To test the hypotheses from our model, we pool the matched dataset and interact the treatment dummies with the pre-park elasticities of demand, factor shares, and other economic

---

[8] The exception is Kerinci Seblat National Park established in 1999. See our prior footnote.

[9] The queen's contiguity based weights matrix defines a location's neighbors as those that either share a border or a vertex. GeoDa is short for Geographic Data Analysis.

characteristics expected to affect leakage. We first use a standard OLS regression, clustering standard errors at the village level, and then a spatial lag regression to control for the fact that deforestation in one parcel is likely affected by deforestation in nearby parcels. We use regional fixed effects to control for unobservable variation in deforestation by region. Some of our variables are correlated because they are generated using local production data, so we use a regression for each interactions alone and also a regression with all interactions together.

We rely on the inherent heterogeneity all across Indonesia to identify the characteristics associated with more or less leakage near the different parks. For example, we compare the effect of a protected area on leakage in areas producing agricultural products with high demand elasticities to areas producing other products with low demand elasticities. The location of protection may be driven by unobservables, but as long as those unobservables are not correlated with the effect of local characteristics on local leakage, then our results will be unbiased. It is not a problem if those unobservables are correlated with the observable economic (or other) characteristics, since we control for those. The scenario that could cause concern is if planners choose to place protected areas in regions where, say, the elasticity of demand for agricultural products in nearby areas is expected to have a larger effect on deforestation than in other regions. In that case, the measured leakage could be associated with specific economic characteristics near the parks rather than triggered by the creation of the park. To explore this possible concern, we use a placebo test to ask whether we observe similar outcomes in other environmentally sensitive areas that did not see a change in protection during our time frame.

### *2.3 Data*

Indonesia covers a total area of 1,904,569 square km and is broadly divided into five island-regions: Java-Bali, Sumatra, Kalimantan, Sulawesi and Papua (as seen in Figure 9). These five regions are then divided into 33 provinces, which are further subdivided into approximately 500 districts.[10] These districts are further subdivided into about 68,000 villages. Our unit of observation for forest cover, protection and biophysical characteristics of land are uniform 3 km by 3 km parcels, for a total of 195,466 parcels for the whole of Indonesia. Since we treat each district as a small separate economy, many of our economic variables are obtained at the district level (i.e. employment in agriculture and wages). However, some of the other economic

[10] The number of districts in Indonesia were 440 in 2000 but increased to 497 in 2010.

variables are estimated at the village level (i.e. price elasticity of demand for agriculture and tree products and share of land in agriculture and tree production).[11]

We use the World Database on Protected Areas (WDPA) to identify parcels of land that are designated as protected, including the six International Union for Conservation of Nature (IUCN) categories: national parks, nature reserves, game reserves, wildlife sanctuaries, recreation parks, and grand forest parks. These conservation areas are state-owned land designated as protected for the conservation of animal and plant species and their ecosystems. They are managed by the Ministry of Forestry, though the decentralization process following the *Reformasi* movement in 1999 has increased the role of local governments in forest conservation projects. All eight protected areas analyzed in this study are designated as national parks.

We use data on annual forest cover change between 2000 and 2012 at the 3 km by 3 km parcel grid size. These data are based on the Hansen *et al.* (2013) study that evaluates global forest cover change from 2000 to 2012. This study improves on some prior datasets by using Landsat data at a finer spatial resolution of only 30 meters. For Indonesia, that study estimates a total forest cover loss of 15 million hectares from 2000 to 2012.

We use Village Potential Statistics based on the 2003 Agriculture census to obtain village-level economic characteristics such as hectares and production quantities devoted to cash crops, plantation products, wood products, forestry products, medicinal plants, and fruits and vegetables. We then allocate these commodities into good $A$ for land-clearing activity and good $T$ for tree-products.

Good $A$ represents the major forest-clearing industries of Indonesia, such as production of cash agricultural crops (cassava, peanuts, soybeans, rice and corn), mining of coal, ores and minerals, and agricultural plantations (e.g. oil palm or coffee). We have data on land use and production quantities associated with these agricultural and plantation activities. We also have regional GDP derived from agricultural and mining activities for 440 districts in Indonesia. We use these data to represent the type of activity within each district that can be categorized as good $A$. Then we obtain demand elasticities from previous literature for Indonesia for 2002 for all products that can be categorized as good $A$ (listed in Table 2). We create a village-level demand

[11] Our decision to use economic variables at the district or village level is partly based on availability of data and partly on what we think is the relevant "economy" within which labor is mobile such that labor and land can be reallocated between production of a particular set of agricultural products and a particular set of tree products.

elasticity for good $A$ by weighting these individual demand elasticities by the production value of each product at the village level. Further, we use the district-level employment in agricultural activities as a share of district-level agricultural GDP to represent the factor share parameter, $\theta_{AL}$. We calculate the share of land in production of $A$ in the village, $\lambda_{RA}$, as the share of land allocated for agricultural activities under the Indonesian 1990 Spatial Plan.

Good $T$ represents any sustainable activity that maintains forest cover, including timber harvested in rotation, collection of forest wood for fuel, medicinal and pharmaceutical products, and carbon sequestration benefits from land that has maintained forest cover but is not under protection. We obtain demand elasticities from previous literature for Indonesia in 2002 for the key timber species found across Indonesia that can be categorized as good $T$ (as seen in Table 2). As for our demand elasticity proxy for good $A$, we create a village-level demand elasticity for good $T$ by weighting these individual demand elasticities by the value of each tree product at the village level. As our measure of production of good $T$, we use variation in the hectares of land for collection of medicinal plants and in production of wood and forestry products (e.g. acacia, bamboo, teak, or mahogany). Also, we estimate the share of land in production of $T$, $\lambda_{RT}$, as the share of land in the village allocated for sustainable tree products, conservation and protection, based on the land zoning provisions under the Indonesia Spatial Plan of 1990.

We also have data on total cash and in-kind wage rate by district. We take the difference between this observed local wage and the urban wage as an indicator of the immobility of labor. Assuming that the nationwide urban wage rate is similar across cities, then we can use the log of the observed district-level wage rate as our proxy for relative immobility (to represent $\sigma_L$, the elasticity of $L_{TA}/L_M$ with respect to $w_M/w$). We also use the district's distance-to-city as an alternative proxy for the immobility of labor between rural and urban sectors.

We recognize that some forest production does not include replanting and is perhaps better reclassified. Most empirical studies of leakage test whether observed forest cover in regions surrounding protected areas rises or falls after the introduction of protection. Our measure of $T$ reflects the fact that even for some unsustainable commercial forest production, we still observe forest cover. Further, as long as the mix of sustainable versus extractive forest production does not vary greatly among our treatment and control parcels, our results on leakage will not be affected.

Under Indonesia's 1990 National Spatial Plan, the forests within each region are divided among three broad land management categories: protection, production, and conversion (Broich *et al.* 2011). Deforestation and logging are strictly prohibited in "protection zones", which include national parks, nature reserves, wildlife sanctuaries, recreational and hunting parks, and watershed protection reserves, as well as some areas outside such parks and reserves. The "production zone" includes areas allocated for commercial selective logging that leads to sustainable forest use (but where deforestation is prohibited). The "conversion zone" includes areas allocated to industrial plantations, smallholder agriculture, mining, urban areas, and government-sponsored transmigration settlements. We use these land allocation categories and protected area boundaries to determine the initial percentages of land allocated to sectors $E$, $T$, and $A$ within each district.[12] Land in *E* includes all parcels that are within protected area boundaries. Land in *T* incudes all parcels in conservation, protection and production zones (outside of protected areas). Land in *A* includes all that are in the conversion zone (outside of protected areas). Table 3 provides details about the area of land in each category based on Indonesia's 1990 National Spatial Plan. We then derive the land share parameters ($\lambda_{RA}$, $\lambda_{RT}$, and $\lambda_{RE}$) for each village, and we determine the amount of land that was taken from each land allocation category to be set aside for protection (to calculate $\alpha$ for each park).

We use the Indonesian Sub-National Growth and Governance dataset to identify the gross regional domestic product (GRDP) and employment shares of major productive sectors of the economy in each district in 2001: agriculture, mining, manufacturing, energy and electricity, construction, trade, transportation, finance, and other services.

For each 3 km by 3 km portion of the grid, we also obtain data on characteristics to include as covariates in the matching process, such as slope, elevation, distance to nearest city, and distance to nearest road. We use ArcGIS[13] and a detailed administrative boundary map of Indonesia to allocate each parcel to a region, province, and district. In Table 4, we summarize the characteristics of parcels that lie within the treatment area and within the control area (as well as characteristics of land within the protected areas that are not used in our regressions). Elevation and slope are highest for parcels within protected areas as compared to parcels outside

[12] The data actually classify land in Indonesia into *six* categories: conservation, protection, production, conversion, non-forest areas, and other. Deforestation and logging are strictly prohibited in the conservation and protection zone, selective logging is permitted in production zone, and deforestation is permitted only in the conversion zone. "Non-forest" represent areas without forest cover, while "other" includes urban areas and all other parcels.

[13] ArcGIS is a geographic information system software for working with maps and information.

protected areas; elevation and slope for parcels within the treatment areas near the new parks are lower than parcels in the control area further away. Parcels within protected areas are closer to cities than parcels outside protected areas. Parcels within the treatment areas are in closer proximity to cities and roads than parcels in the control areas.

In Figure 10, we see average forest cover for parcels in the areas near the new parks and areas in the matched counterfactuals for each year from 2000 through 2012. The figure indicates that the trend in forest cover is similar across treatment and matched parcels before the seven new parks were established in 2004.[14]

Table 5 provides summary statistics for some of the key economic variables of the matched dataset. Both price elasticities of demand for good *A* and good *T* are similar across treatment areas and matched controls. Only 5% of the total land taken to establish the eight new protected areas is from land designated for agriculture; the remaining 95% of the set-aside is taken from land designated for sustainable tree products. The initial share of total land designated for sustainable tree products ($\lambda_{RT}$) is similar across treatment and matched control parcels, but the share of land designated for agriculture ($\lambda_{RA}$) is higher within treatment parcels. The ratio of agricultural workers to district-level gross regional domestic product (GRDP) from agriculture ($\theta_{AL}$) is lower within treatment areas. The wage rate is similar between treatment and matched control parcels.

## 2. Results

Our first hypotheses states that leakage will be negative if and only if $\alpha < \frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)}$. To test this hypothesis, we calibrate the model to the area around each new park. Then we use each park's parameter values to calculate that ratio, the direction of the inequality, and the predicted sign of leakage. We then measure actual leakage around each park to see if it has the predicted sign.

As shown in Table 6, these calibrations predict that leakage will be negative in areas immediately around six of the new parks and positive around the other two. To measure actual leakage, we compare treated parcels to other counterfactual parcels using the nearest neighbor matching process. In fact, the counterfactuals for this leakage calculation are the same control

[14] We cannot compare this trend for the treatment parcels near the protected area that was established in 1999 (i.e. Kerinci Seblat National Park), because we only have forest cover data from 2000.

parcels as used in regressions below. Here, we estimate the average treatment effect on the treated (ATT) for forest loss in parcels near each of the eight newly established protected areas compared to these controls; this ATT serves as our estimate of observed leakage.

Results in Table 6 show statistically significant positive leakage near two parks, Aketajawe Lolobata National Park (+7.4%) and Tesso Nilo National Park (+6.9%). In areas near the other six parks, we find evidence of negative leakage ranging from –0.2% to –10.3%. As seen in Table 6, our theoretical model and calibration correctly predicts positive leakage around one of the two parks found to have positive leakage. And it correctly predicts negative leakage for five of the six parks found to have negative leakage. Overall, the sign on predicted leakage matches the sign on observed leakage for six of the eight national parks. Five of those estimates are statistically significant in the predicted direction, and two are statistically significant in the opposite direction. Thus, while most of these results are consistent with our theoretical model, the first hypothesis is not unambiguously proven using only these eight observations.[15]

In Table 7a, we assess the quality of the matches used for estimating leakage by comparing summary statistics of the covariates for control and treatment parcels that were used in the nearest neighbor matching process. The fact that the normalized differences of covariates used in the matching process are less than 0.25 standard deviations suggests that the matched control parcels are similar to the treatment parcels (Imbens and Wooldridge, 2008). In Table 7b, we also check covariate balances for parcel-level characteristics that are not included in the matching process, such as distance to river, spatially lagged values of distance to river, and accessibility. For these characteristics, the normalized differences between treatment parcels and matched control parcels are also lower than 0.25, further indicating that the matching process is successful in finding control parcels that are similar to treatment observations.

Next, we use the same matched dataset for empirical tests of hypotheses 2 through 8. We test these hypotheses in two ways. First, we run separate regressions of forest cover loss on covariates described above and on one of the key economic variables at a time. In all cases, the empirical test of our hypothesis depends on the significance of the sign of the estimated coefficient on the interaction between the treatment dummy and each key economic variable. For each of those separate regressions, a row of Table 8 shows the estimated coefficient only for the

[15] With higher assumed values of 0.5 for elasticities of substitution in production of *A* and *T*, then the first theorem would lead to an hypothesis that all eight parks have negative leakage. In this case, the model would still yield six correct predictions for the eight parks.

key interaction term, both for the OLS and the IV spatial lag regression. For a second way of testing these hypotheses, we run a regression of forest cover loss on the covariates described above and all the key economic variables simultaneously in one regression (omitting only the share of land in *T*).[16] Table 9 shows results from this latter regression. The spatial lag in Table 9 is positive and significant in the IV regression, indicating that deforestation in one parcel leads to an increase in deforestation in neighboring parcels. All regressions include treated parcels in the 35 districts contiguous to the new parks and matched parcels outside these districts.

Our second hypothesis states that areas where more of the new park is taken away from designated agricultural land will experience more leakage, i.e., forest will be cut to replace some of that taken agricultural land. Our empirical results support this hypothesis, as seen by the positive coefficient on the share of land taken from *A* in both tables, although this coefficient is significantly different from zero only in the spatial lag regressions.

Our third hypothesis is that leakage will be smaller (or more negative) if the demand for agricultural products is more elastic. The removal of land raises the equilibrium price of those products, with less need to replace that land if demand is elastic. In Tables 8 and 9, we find evidence supporting this hypothesis. The coefficient for the price elasticity of demand for good *A* interacted with the treatment dummy is negative, and significant in the OLS regressions. The indication is that parcels near the new parks that produce good *A* with inelastic demand tend to have more deforestation than areas producing agricultural products with more elastic demand.

Our fourth hypothesis posits that leakage will be larger if the demand for the tree products produced in that area is more elastic (because production of tree products can shrink). Indeed, the coefficient on the interaction term between the treatment dummy and the price elasticity for good *T* is positive and significant in both tables (in both OLS and IV). Thus, we find that the empirical results corroborate our expectation of an increase in deforestation in nearby areas that have a higher price elasticity of demand for their good *T*.

Our fifth hypothesis posits that more land initially allocated for agriculture will lead to less leakage (because any taking of agricultural land for the new protected area can be more easily replaced by the large existing agricultural land base). Our estimated coefficient on the share of land in agriculture interacted with nearby areas is negative as predicted (in both columns

[16] We omit the share of land in *T* from the full regression because it is highly collinear with the share of land in *A*.

of both tables), but the estimate is not significantly different from zero.

Hypothesis six says that a greater initial share of land allocated to timber production will lead to larger leakage.[17] When we regress forest loss on that initial share of land for timber interacted with the indicator variable for nearby parcels in Table 8, the coefficient is positive and significant. Thus, when more land is already used for production of good $T$, all else equal, nearby areas experience greater deforestation after the park is established. Essentially, in that case, more of that existing timber land is available for reallocating to good $A$.

Our seventh hypothesis is that a higher factor share for labor in agriculture ($\theta_{LA}$) will be associated with smaller leakage. We proxy for this factor share using district-level employment in agricultural activities as a share of district-level agricultural GDP. As predicted, Tables 8 and 9 show that this key economic variable interacted with the treatment dummy has a coefficient that is always negative (and usually significant). The intuition from the theoretical model is that the taking of agricultural land for a new park is not as critical for areas that produce labor-intensive agricultural goods (as opposed to land-intensive agricultural goods).

Our last hypothesis posits that when labor is more mobile from the rural sectors (timber and agriculture) to urban manufacturing ($M$), then forest leakage will be smaller (because agricultural labor can move to find employment elsewhere, with less need to cut trees for land to employ those workers in the area near the new park). To proxy for this mobility, we use the wage rate in the district near a new park (before the park is established). Wage rates in rural areas tend to be less than in urban areas, so a lower wage in a particular rural area implies less labor mobility. That is, greater mobility of labor will tend to be associated with a higher wage in the rural area compared to its urban counterpart. Indeed, as predicted, Tables 8 and 9 show that a higher local rural wage is associated with less leakage after the park is established.

## 4. Conclusion

In this paper, we build a simple theoretical general equilibrium model that can predict positive or negative leakage in a way that depends on key economic characteristics of the local forest goods and non-forest goods. For protected areas in Indonesia established between 1999 and 2004, compare nearby areas to matched controls further away. We use data on the

[17] This hypothesis is similar to the previous one, but it is not identical because the shares of land for tree products and for ag products do not add to one. Because many or most districts contain only agricultural and timber land, however, we cannot include both of these shares in the single regression of Table 9.

differences in product demand and production processes near each park to estimate whether nearby leakage is driven by differing market factors. Our model helps predict where and when leakage is most likely to affect the success of protecting forest, allowing policy-makers either to target their conservation efforts or to find alternative methods for mitigating leakage.

Based on the general equilibrium model, we expect leakage to be smaller or negative in cases with a larger initial allocation of land for agricultural production, or when the price elasticity of demand for such sustainable forestry products is small. Following the policy shock, markets may reallocate land from agricultural and extractive industries to these forest industries in order to maintain production of forest products. Conversely, when the price elasticity of demand for agricultural output is small, or when the initial land allocated for sustainable tree products is large, the model predicts larger leakage, as the markets may reallocate land from production of sustainable forestry goods to production of agricultural goods under such scenarios. Leakage is also expected to be larger when more land is taken away from agricultural production for new protection and when the factor share of land in agriculture is larger.

Our empirical results confirm almost all of these hypotheses. We first find that the "nearby" areas with larger share of land taken from extractive industries to create new protected areas experience an increase in deforestation pressures. Thus, policy-makers may achieve higher success from protection policy if they target areas that are designated for sustainable tree products rather than agricultural production.

Next, we find that areas near new parks experience lower deforestation pressures if they face higher price elasticity of demand for their agricultural outputs. Thus, future policy decisions could designate new protected areas in regions that have higher price elasticity of demand for agricultural and extractive industries. To reduce the potential increase in deforestation levels following the policy shock, policy-makers may also want to consider incentive-based payments or greater enforcement in areas near recently established protected areas with lower price elasticity of demand for their agricultural output.

Empirical results also show that nearby areas with a larger initial allocation of land for ag and extractive industries can face lower deforestation pressures. The policy shock may lead to smaller relative reduction of this sector, and hence some land may be reallocated to sustainable forestry production. Future conservation policy can focus more on areas where large amount of resources such as land are allocated for agriculture and extracion, such that a policy shock that reduces these resources for conservation will not lead to major reallocation from sustainable sectors to extractive sectors.

Empirically, we find relatively higher deforestation pressures in nearby areas with larger employment in extractive industries. A policy implication is that areas with a large proportion of people employed in agriculture and mining activities could be more stringently monitored following establishment of protected areas in order to constrain leakage. These areas can also benefit from an increase in non-extractive industries that are labor intensive,to provide employment opportunities for workers following the establishment of new protected areas.

Finally, we find that areas with greater labor mobility have less forest leakage Thus, one option to limit leakage may be to facilitate labor movement out of local resource-extracting industries.

**Table 1: National Average Parameter Values for Simulation Exercises**

| Variable Name | Variable Symbol | Value |
|---|---|---|
| Initial fraction of land allocated to good *E* | $\lambda_{RE}$ | 0.04 |
| Initial fraction of land allocated to good *A* | $\lambda_{RA}$ | 0.28 |
| Initial fraction of land allocated to good *T* | $\lambda_{RT}$ | 0.68 |
| Share of $dR_E$ taken from land initially allocated for production of good *A* | $\alpha$ | 0.05 |
| Demand elasticity for good *T* | $\varepsilon$ | -0.50 |
| Demand elasticity for good *A* | $\eta$ | -0.49 |
| For good *A*, labor share of production [a] | $\theta_{AL}$ | 0.40 |
| For good *A*, land share of production [a] | $\theta_{AR}$ | 0.60 |
| For good *T*, labor share of production [b] | $\theta_{TL}$ | 0.10 |
| For good *T*, land share of production [b] | $\theta_{TR}$ | 0.90 |
| Factor share of labor in good *T* | $\lambda_{LT}$ | 0.15 |
| Factor share of labor in good *A* | $\lambda_{LA}$ | 0.85 |
| Factor share of labor in manufacturing [c] | $\lambda_{LM}$ | 1.00 |
| Elasticity of substitution between *L* and *R* in production of the agricultural good [d] | $\sigma_A$ | 0.20 |
| Elasticity of substitution between *L* and *R* in production of tree products [e] | $\sigma_T$ | 0.20 |
| Elasticity of substitution between rural labor ( $L_{TA}$) and urban labor ($L_M$) [f] | $\sigma_L$ | 0.30 |

[a,b,c,d,e,f] Due to unavailability of data, values for these parameters are assumed.

**Table 2: Elasticities for Agricultural Output and Sustainable Tree Products**

| Agricultural Products | Elasticity | Tree Products | Elasticity |
|---|---|---|---|
| Fruits | -0.95 | Teak | -0.05 |
| Vegetables | -1.11 | Meranti | -0.20 |
| Plantation Products (includes palm oil) | -0.40 | Mahogany | -0.10 |
| Cassava | -0.33 | Pine | -0.80 |
| Maize | -0.82 | Camphor | -0.50 |
| Rice | -0.42 | Keruing | -0.50 |
| Soybean | -0.54 | Acacia | -0.10 |
| Sweet Potatoes | -0.40 | Bamboo | -0.60 |
| Peanuts | -0.41 | Eucalyptus | -0.20 |
| | | Walnuts | -0.10 |
| | | Lamtoro | -0.90 |
| | | Rattan | -1.00 |
| | | Sengon | -0.60 |
| | | Sono | -0.50 |
| | | Trembesi | -0.50 |

Sources: Deaton 1990; Adamovicz and Dyrcz 2008; Chimeli and Soares 2011; Abildtrup *et al.* 2012; Dermoredjo *et al.* 2013.

**Table 3: Land Uses in Indonesia and our Allocation in the Data**

| Name of Category in the 1990 National Spatial Plan of Indonesia | Activities Included | Allocated to | Total Area in Indonesia (sq. km) | % |
|---|---|---|---|---|
| Prior Protected Areas | Protected areas established before 1999 | *E* | 190,883 | 11.00% |
| New Protected Areas | Protected areas established between 1999 and 2004 | *dE* | 27,512 | 1.58% |
| **Areas Outside of Protected Areas** | | | | |
| Protection | Deforestation and logging strictly prohibited | *T* | 409,864 | 23.61% |
| Conservation | Deforestation and logging strictly prohibited | *T* | 21,385 | 1.23% |
| Production | Deforestation prohibited but selective logging allowed | *T* | 461,627 | 26.59% |
| Conversion | Deforestation is permitted | *A* | 191,343 | 11.02% |
| Non-Forest | Areas not considered forest | N/A | 431,903 | 24.88% |
| Other | Areas not in above categories | N/A | 1,373 | 0.08% |
| TOTAL | | | 1,735,890 | 100% |

**Table 4: Summary Statistics for Parcels within Protected Areas, within Treatment Areas, and Outside Treatment Areas**

| | **New Protected Areas** | | **Treated Parcels in Nearby Districts** | | **Parcels outside Nearby Districts** | |
|---|---|---|---|---|---|---|
| **Variable** | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Neighbors' Average [a] Forest Cover Loss 2000-2012 (ha) | 43.04 | 103.09 | 153.88 | 171.20 | 73.06 | 121.18 |
| Forest Cover Loss 2000-2012 (ha) | 42.66 | 122.41 | 156.39 | 210.42 | 74.52 | 146.05 |
| Neighbors' Average Pre-Trend [b] (ha) | 4.13 | 14.04 | 17.66 | 33.45 | 6.70 | 14.15 |
| Pre-Trend [b] (ha) | 4.16 | 19.33 | 17.94 | 50.00 | 6.86 | 21.70 |
| Neighbors' Average Distance to City (km) | 74.40 | 75.88 | 89.69 | 73.98 | 126.98 | 99.89 |
| Distance to City (km) | 74.68 | 76.81 | 92.62 | 79.11 | 129.83 | 100.80 |
| Neighbors' Average Distance to Road (km) | 31.10 | 29.03 | 26.38 | 36.94 | 60.82 | 84.05 |
| Distance to Road (km) | 31.23 | 29.10 | 26.87 | 37.28 | 62.71 | 87.16 |
| Neighbors' Average Elevation (m) | 702.15 | 549.64 | 207.82 | 290.35 | 312.34 | 455.14 |
| Elevation (m) | 712.32 | 581.18 | 206.89 | 299.90 | 315.01 | 490.46 |
| Neighbors' Average Slope (deg) | 11.75 | 7.07 | 5.24 | 5.10 | 6.92 | 6.39 |
| Slope (deg) | 11.94 | 7.66 | 5.26 | 5.53 | 7.02 | 6.86 |
| No. of Observations | 3057 | | 17088 | | 153811 | |

[a] Neighbors are defined as those polygons that share a continuous boundary (where we assume that these values from neighboring polygons can be weighted equally).

[b] "Pre-trend" refers to the average forest cover loss from 2000 to 2002.

Notes: (1.) Each parcel is 900 hectares; for example, a mean forest cover loss of 43 for new protected areas means that the total forest cover loss from 2000 to 2012 for these parcels averaged 4.8%. The above table represents summary statistics for the parcels remaining after we drop all parcels that were protected before 1999 and parcels with zero forest cover in 2000.

**Table 5: Summary Statistics for Key Economic Variables in Matched Dataset**

| | Treatment Parcels | | Matched Counterfactuals | |
|---|---|---|---|---|
| **Variable** | Mean | Std. Dev. | Mean | Std. Dev. |
| Share of new park taken from $A$ ($\alpha$) | 0.05 | 0.16 | 0.00 | 0.00 |
| Elasticity of demand for good $A$ ($\eta$) | 0.49 | 0.13 | 0.50 | 0.13 |
| Elasticity of demand for good $T$ ($\varepsilon$) | 0.50 | 0.36 | 0.48 | 0.34 |
| Initial fraction of land in $T$ ($\lambda_{RT}$) | 0.53 | 0.42 | 0.55 | 0.44 |
| Initial fraction of land in $A$ ($\lambda_{RA}$) | 0.22 | 0.33 | 0.15 | 0.30 |
| Ag employment/Ag GRDP [a] ($\theta_{AL}$) | 1.46 | 0.87 | 1.75 | 0.88 |
| Log of wage rate [b] ($\sigma_L$) | 13.20 | 0.25 | 13.17 | 0.25 |

[a] This ratio represents the number of people employed in agriculture at the district level divided by the district-level GRDP for agriculture.

[b] The wage rate represents the ratio of total wages (i.e. cash plus in-kind) at the district-level divided by the number of people employed in that district.

**Table 6: Predictions and Estimates of Leakage Near Eight New National Parks**

| National Park | Leakage (ha) (std. errors in parentheses) | Leakage (%) | Sign of Predicted Leakage | Does the Model Predict Correct Direction? |
|---|---|---|---|---|
| Aketajawe Lolobata | 67.01 (4.95) | 7.45% | – | N |
| Bantimurung Bulusaraung | -28.59 (2.04) | -3.18% | – | Y |
| Batang Gadis | -4.32 (5.79) | -0.48% | – | Y |
| Gunung Ciremai | -2.66 (1.22) | -0.30% | – | Y |
| Gunung Merbabu | -1.78 (0.85) | -0.20% | – | Y |
| Kerinci Seblat | -9.56 (2.49) | -1.06% | + | N |
| Sebangau | -92.78 (6.91) | -10.31% | – | Y |
| Tesso Nilo | 61.96 (5.5) | 6.88% | + | Y |

Notes: Calculations for these model predictions assume $\sigma_T = \sigma_A = 0.2$ and $\sigma_L = 0.3$.

**Table 7a: Covariate Balance for Variables Used in the Matching Process**

| | Treatment Parcels | | Control Parcels | | |
|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Norm. Diff.** [b] |
| Neighbors' Average [a] Forest Cover 2000 (ha) | 17.66 | 33.45 | 16.76 | 31.58 | 0.02 |
| Forest Cover 2000 (ha) | 17.94 | 50.00 | 16.79 | 47.65 | 0.02 |
| Neighbors' Average Forest Loss 2000-2012 (ha) | 153.87 | 171.20 | 157.16 | 174.82 | 0.01 |
| Forest Loss 2000-2012 (ha) | 156.38 | 210.40 | 159.15 | 210.54 | 0.01 |
| Neighbors' Average Distance to City (km) | 89.68 | 73.98 | 87.29 | 80.68 | 0.02 |
| Distance to City (km) | 92.61 | 79.11 | 90.26 | 86.35 | 0.02 |
| Neighbors' Average Distance to Road (km) | 26.38 | 36.94 | 28.98 | 38.47 | 0.05 |
| Distance to Road (km) | 26.86 | 37.28 | 29.93 | 39.82 | 0.06 |
| Neighbors' Average Elevation (m) | 207.80 | 290.33 | 192.66 | 292.14 | 0.04 |
| Elevation (m) | 206.88 | 299.88 | 190.51 | 299.04 | 0.04 |
| Neighbors' Average Slope (deg) | 5.24 | 5.10 | 5.11 | 5.07 | 0.02 |
| Slope (deg) | 5.26 | 5.53 | 5.14 | 5.48 | 0.02 |

[a] Neighbors are defined as those polygons that share a continuous boundary (where we assume that these values from neighboring polygons can be weighted equally).

[b] The normalized difference is the ratio of the difference in the average covariate value between treatment and control parcel to the square root of the sum of the squared standard deviations of treatment and control parcels.

**Table 7b: Covariate Balance for Variables Not Used in the Matching Process**

| | Treatment Parcels | | Control Parcels | | |
|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Norm. Diff.** [c] |
| Neighbors' Average [a] Distance to River (m) | 3842.63 | 4007.65 | 3610.69 | 3540.20 | 0.04 |
| Distance to River (m) | 3756.52 | 3546.55 | 3542.65 | 3112.88 | 0.05 |
| Accessibility [b] (minutes) | 678.57 | 811.96 | 591.63 | 733.63 | 0.08 |

[a] Neighbors are defined as those polygons that share a continuous boundary (where we assume that these values from neighboring polygons can be weighted equally).

[b] "Accessibility" is provided in these data as the average travel time in minutes to the nearest city with 50,000 inhabitant in 2000.

[c] The normalized difference represents the ratio of the difference in the average covariate value between treatment and control parcel to the square root of the sum of the squared standard deviations of treatment and control parcels.

**Table 8: Key Coefficients from Individual Regressions of Forest Loss on Covariates and One Key Variable at a Time**

| Economic Parameter from Theory Model | Expected Sign from Theory Model | Proxy Economic Variable | Coefficient Estimate from Individual OLS Regression (clustered standard errors) | Marginal Effect from Individual IV Spatial Lag Regressions (robust standard errors) |
|---|---|---|---|---|
| Share of park taken from *A* ($\alpha$) | + | Treatment × Share of park taken from *A* | 33.39<br>(34.51) | 193.99**<br>(93.14) |
| Elasticity of demand for *A* ($\eta$) | - | Treatment × Elasticity of *A* | -75.58***<br>(18.53) | -33.05<br>(46.43) |
| Elasiticity of demand for *T* ($\varepsilon$) | + | Treatment × Elasticity of *T* | 28.31***<br>(5.34) | 39.09***<br>(13.13) |
| Initial fraction of land in *A* ($\lambda_{RA}$) | - | Treatment × Fraction of land in *A* | -0.95<br>(8.96) | -14.59<br>(23.94) |
| Initial fraction of land in *T* ($\lambda_{RT}$) | + | Treatment × Fraction of land in *T* | 75.39***<br>(3.58) | 127.97***<br>(13.39) |
| Factor share of labor in agriculture ($\theta_{AL}$) | - | Treatment × Ratio of ag. workers /ag. output | -86.28***<br>(28.00) | -66.54<br>(73.55) |
| Labor mobility between rural (*A* and *T*) vs urban (*M*) sector ($\sigma_L$) | - | Treatment × Log of wage rate | -50.30***<br>(8.56) | -40.56*<br>(22.23) |

*** p<0.01, ** p<0.05, * p<0.1

Note: The OLS column shows the estimated coefficient (which for OLS is also the marginal effect).

**Table 9: Regressions of Forest Loss on All Variables Together (t-statistic in parentheses)**

| | **OLS** | **Instrumental Variable** |
|---|---|---|
| Spatial lag in forest cover loss 2000-2012 | | 0.880*** |
| | | (0.0143) |
| Treatment dummy | 852.6*** | 611.63* |
| | (127.4) | (345.91) |
| Distance to city (km) | 0.168*** | -0.06 |
| | (0.0398) | (0.12) |
| Distance to road (km) | 0.458*** | 0.99*** |
| | (0.0773) | (0.19) |
| Elevation (m) | -0.0804*** | -0.08*** |
| | (0.00349) | (0.01) |
| Slope (deg) | -7.395*** | -10.16*** |
| | (0.219) | (0.66) |
| Elasticity of demand for *A* | 33.05** | -4.65 |
| | (15.08) | (32.39) |
| Elasticity of demand for *T* | -15.90*** | -24.07** |
| | (5.114) | (12.20) |
| Fraction of total land initially in *A* | 1.335 | -6.46 |
| | (7.991) | (18.52) |
| Ratio of employment in *A* to district-level GRDP from *A* | -23.30*** | -35.45*** |
| | (2.101) | (5.08) |
| Log of wage rate | 80.07*** | 75.35*** |
| | (7.890) | (18.50) |
| Treatment × $\alpha$ (Share of park taken from *A*) | 36.78 | 191.71** |
| | (33.70) | (92.85) |
| Treatment × $\eta$ (Elasticity of demand for *A*) | -85.48*** | -49.30 |
| | (18.75) | (45.85) |
| Treatment × $\varepsilon$ (Elasticity of demand for *T*) | 22.81*** | 32.86** |
| | (5.643) | (14.48) |
| Treatment × $\lambda_{RA}$ (Fraction of land in *A*) | -3.770 | -18.76 |
| | (9.619) | (24.95) |
| Treatment × $\theta_{LA}$ (Ratio of employment in *A* to district-level GRDP of *A*) | -177.7*** | -149.58* |
| | (30.18) | (80.19) |
| Treatment × $\sigma_L$ (Log of wage rate) | -62.45*** | -46.15* |
| | (9.587) | (26.30) |
| **Regional FE** | yes | yes |
| **Observations** | 90,446 | 90,446 |
| **R-squared** | 0.223 | 0.739 |

*** p<0.01, ** p<0.05, * p<0.1

Notes: (1.) The OLS column shows the estimated coefficient (which for OLS is also the marginal effect). The IV column shows the coefficent for the spatial lag and the marginal effects for all other variables.

(2.) The parameter $\boldsymbol{\alpha}$ appears only with the interactin, because it is only relevant for treatment parcels; it is always zero for control parcels.

(3.) We drop the fraction of land in *T* (and treatment × fraction of land in *T*) from the full regression because of high collinearity between this variable and the fraction of land in *A*.

**Figure 1: Schematic Representation of our Leakage Definition**

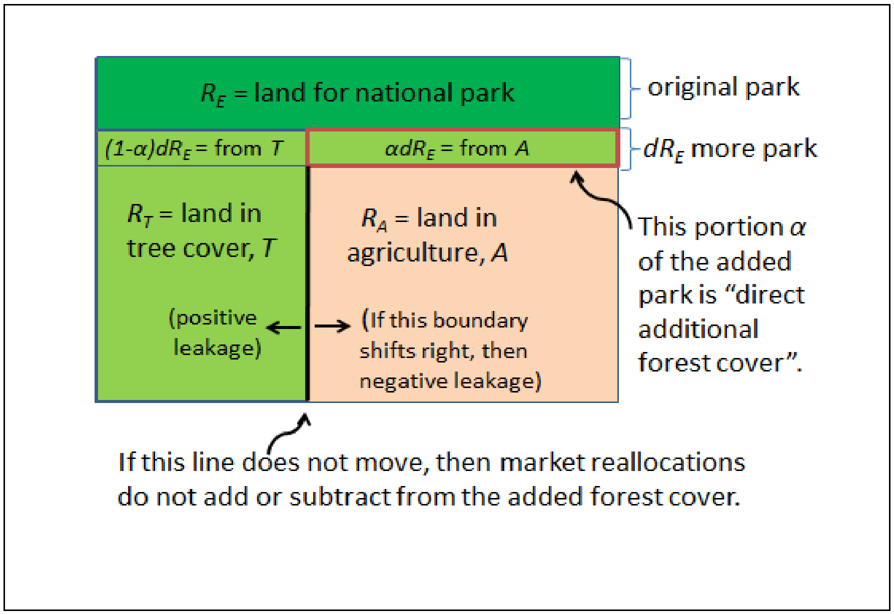
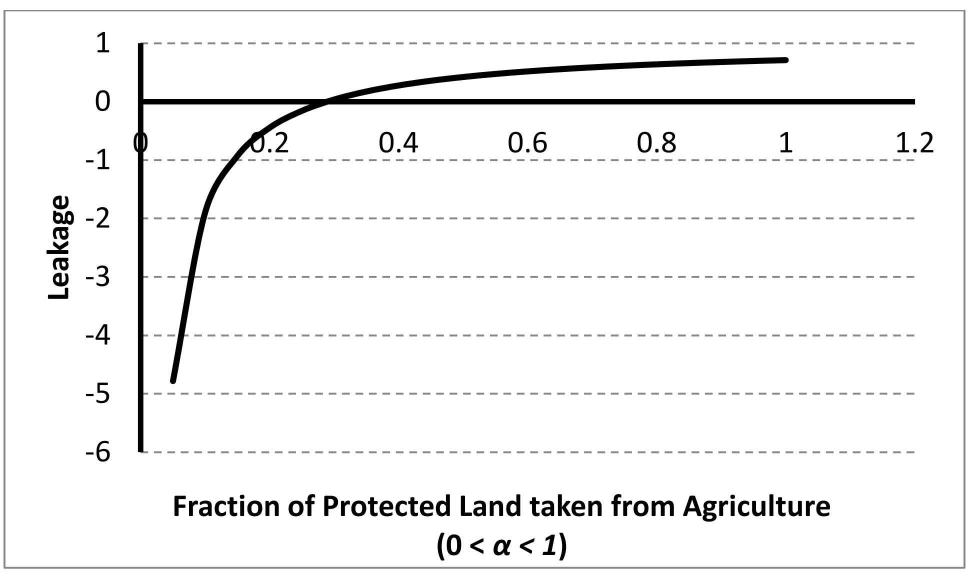
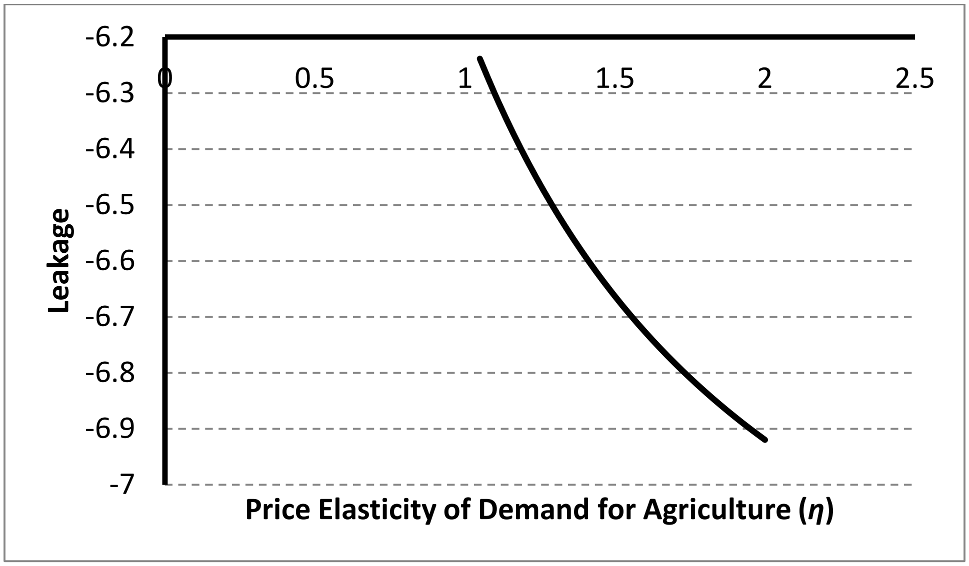



**Figure 2: Effect on Leakage of Variation in the Fraction of the Set-Aside taken from Agriculture**



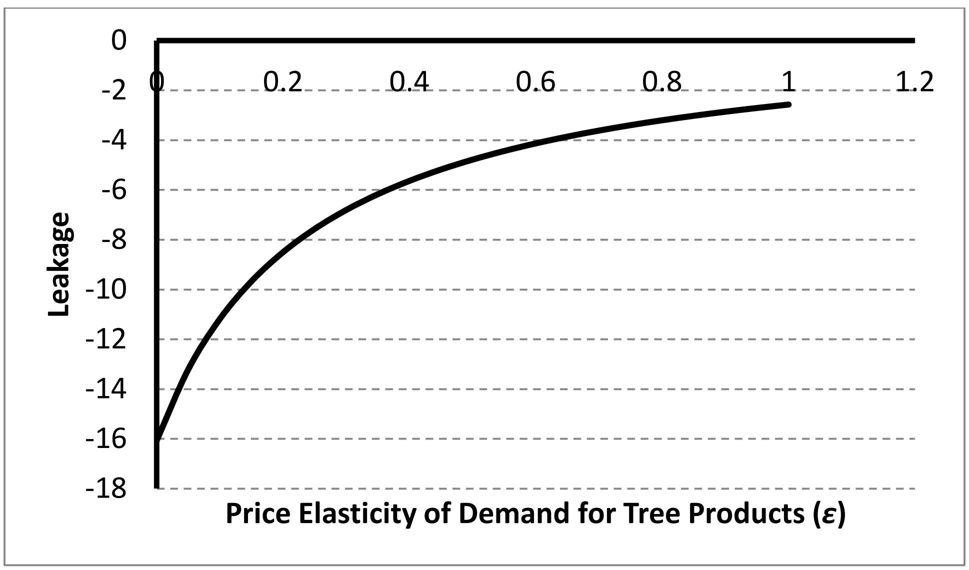

Note: The values used for the figure are shown in Table 1.

**Figure 3: Effect on Leakage of Variation in the Price Elasticity of Demand of Good *A***



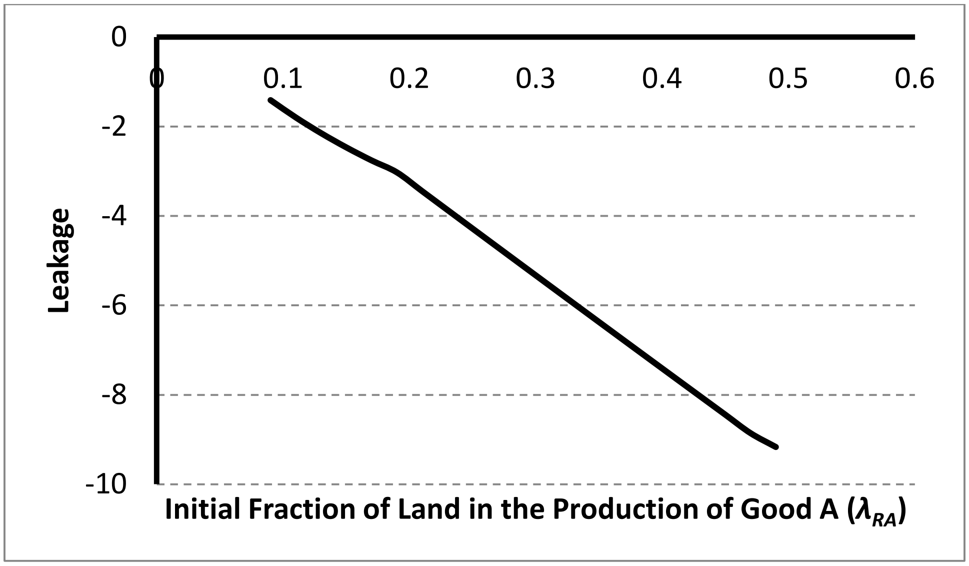

Note: The values used for the figure are shown in Table 1.

**Figure 4: Effect on Leakage of Variation in the Price Elasticity of Demand of Good *T***



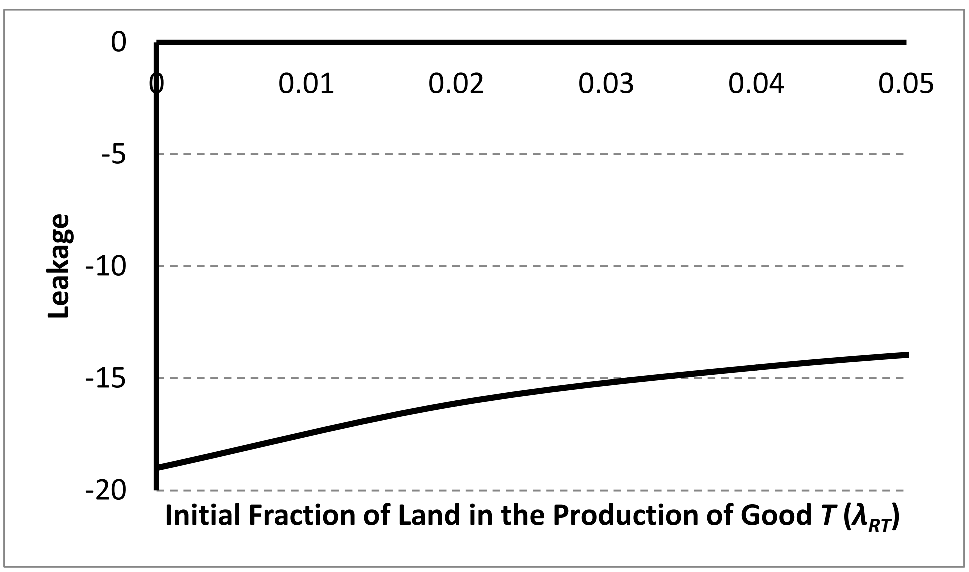

Note: The values used for the figure are shown in Table 1.

**Figure 5: Effect on Leakage of Variation in the Initial Allocation of Resource *R* to Good *A***



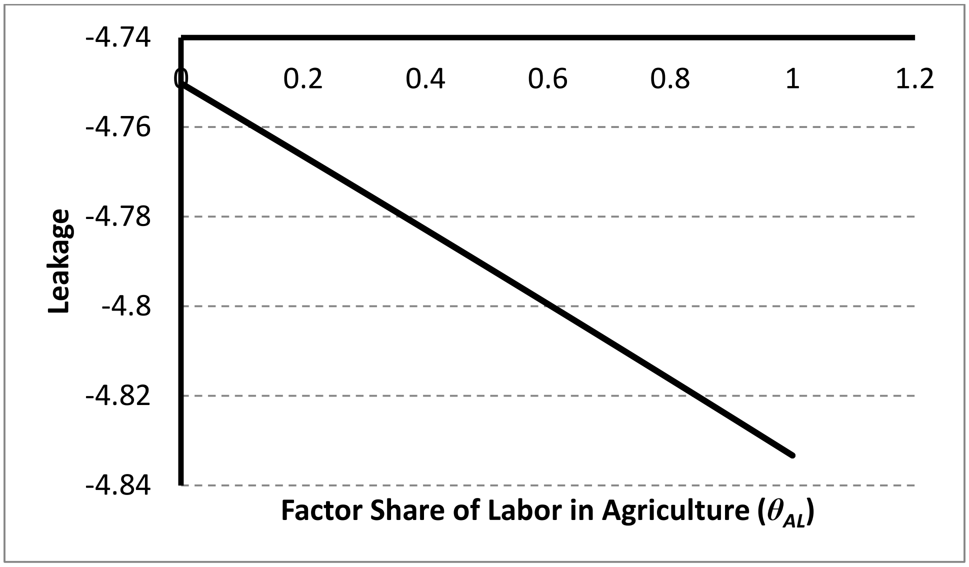

Note: The values used for the figure are shown in Table 1.

**Figure 6: Effect on Leakage of Variation in the Initial Allocation of Resource *R* to Good *T***



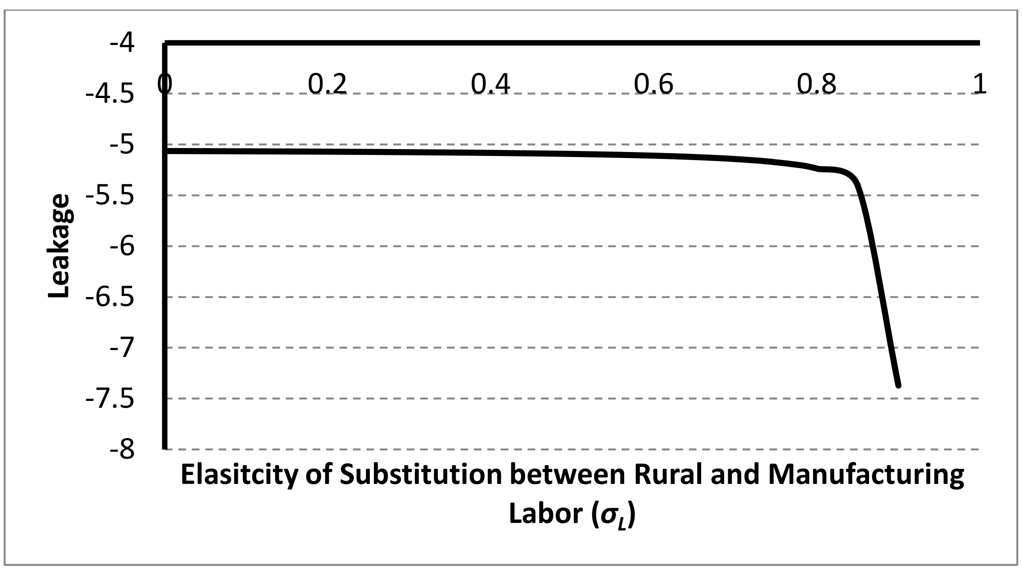

Note: The values used for the figure are shown in Table 1.

**Figure 7: Effect on Leakage of Variation in the Factor Share of Labor in the Production of Good *A***




Note: The values used for the figure are shown in Table 1.

**Figure 8: Effect on Leakage of Variation in the Elasticity of Subsitution between Rural and Manufacturing Labor**




Note: The values used for the figure are shown in Table 1.

**Figure 9: Protected Areas Established between 1999 and 2004, and Our Treatment Regions for Indonesia**

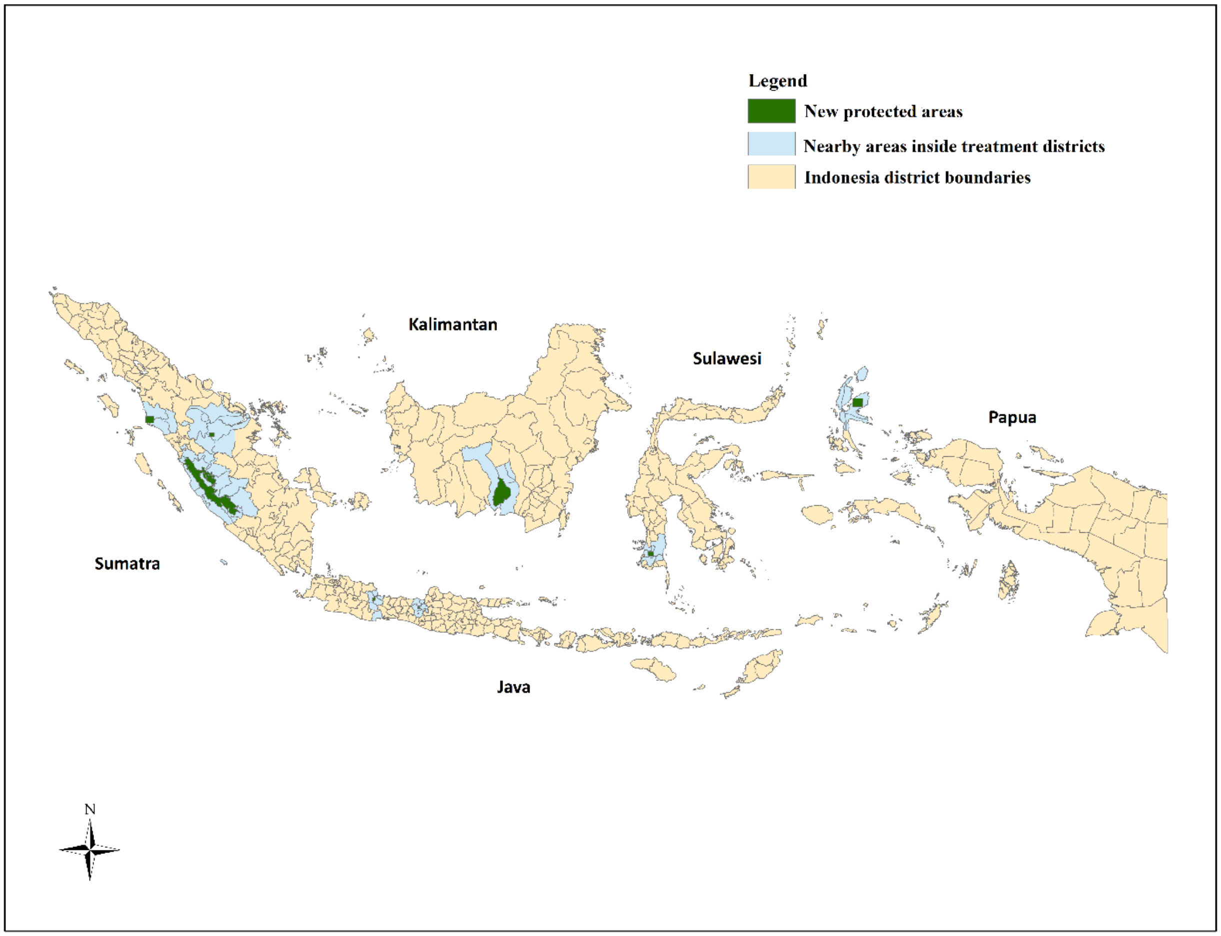
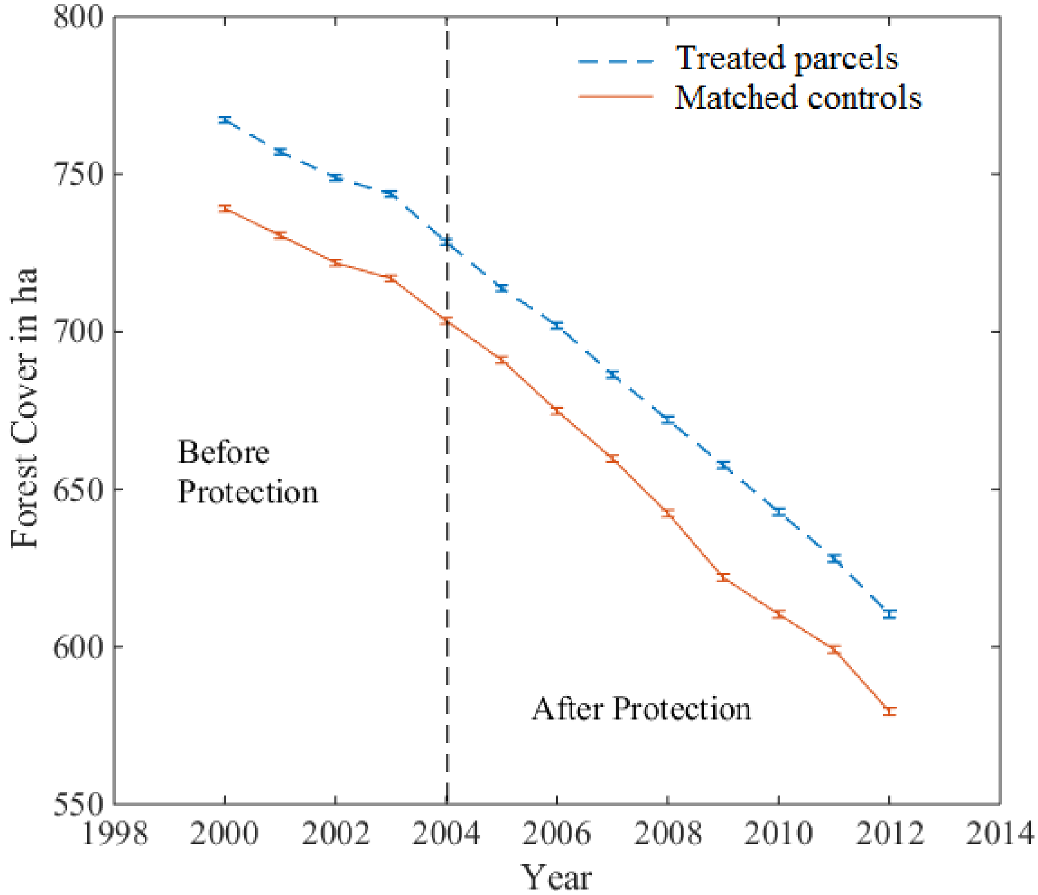


**Figure 10: Forest Cover Near the Seven National Parks and Counterfactuals**




Note: This figure illustrates the average forest cover for each year from 2000 to 2012 for treatment parcels in the districts near the new parks and for the matched control parcels.

## References

Abadie, A., and G. W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica,* 74**,** 235-267.

Abildtrup, J., Helles, F., Holten-Andersen, P., Larsen, J. F., and Thorsen, B. J. (Eds.). 2012. *Modern time series analysis in forest products markets* (Vol. 58). Springer Science and Business Media.

Adamowicz, K., and Dyrcz, A. 2008. An attempt to assess price elasticity of demand for pine wood on the primary wood market in the Bytnica Forest Division in the years 1997-2005. *Acta Scientiarum Polonorum Silvarum Colendarum Ratio et Industria Lignaria*, 7(3), 5-13.

Alix-Garcia, J. 2007. A spatial analysis of common property deforestation. *Journal of Environmental Economics and Management,* 53(2), 141-157.

Alix-Garcia, J., McIntosh, C., Sims, K., and Welch, J. 2010. *Development and Deforestation in Mexico: Impacts Using the Discontinuity in Eligibility for Oportunidades*. Working paper, Department of Agricultural and Applied Economics, University of Wisconsin, Madison.

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa G. A., and Robalino J. A. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences,* 105, 16089–16094.

Angelsen, A. and Kaimowitz, D. 1999. Rethinking the causes of deforestation: Lessons from economic models. *The World Bank Research Observer*, 14(1), 73-98.

Anselin, L. 2002. Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247-267.

Armsworth, P.R., Gretchen, D. C., Kareiva, P. and Sanchirico, J.N. 2006. Land market feedbacks can undermine biodiversity conservation. *Proceedings of the National Academy of Sciences,* 103(14), 5403-5408.

Bode, M., Tulloch, A. I., Mills, M., Venter, O., and W Ando, A. 2015. A conservation planning approach to mitigate the impacts of leakage from protected area networks. *Conservation Biology,* 29(3), 765-774.

Brock, W. A., and Durlauf, S. N. 2001. Interactions-based models. *Handbook of Econometrics*, 5, 3297-3380.

Broich, M., Hansen, M. C., Potapov, P., Adusei, B., Lindquist, E., and Stehman, S. V. 2011. Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation*, 13(2), 277-291.

Brokington, D. and Igoe, J. 2006. Eviction for conservation: A global overview. *Conservation and Society,* 4, 424-470.

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., and Sieber, S. 2011. *The political economy of deforestation in the tropics* (No. w17417). National Bureau of Economic Research.

Butler, R. 2013. World's largest REDD project finally approved in Indonesia. *The Guardian*. May 31.

Chimeli, A., and Soares, R. R. 2011. The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon. MIMEO, Universidade Catolica de Rio.

Curran, L.M., Trigg, S. N., McDonald, A. K., Astiani, D., Hardino, Y. M., Siregar, P., Caniago, I., and Kasischke, E. 2004. Lowland forest loss in protected areas of Indonesian Borneo. *Science*, 303, 1000–1003.

Deaton, A. 1990. Price elasticities from survey data: extensions and Indonesian results. *Journal of econometrics*, 44(3), 281-309.

Deininger, K., and Minten, B. 2002. Determinants of deforestation and the economics of protection: An application to Mexico. *American Journal of Agricultural Economics*, 84(4), 943-960.

Dermoredjo, S. K., Sudaryanto, T. and Masyhuri. 2013. The impacts of AFTA to the main food crops sector in Indonesia. Food and Fertilizer Technology Center (FFTC) Agricultural Database. URL: http://ap.fftc.agnet.org/files/ap_policy/108/108_1.pdf

Ferraro, P.J. 2002. The local costs of establishing protected areas in low-income nations: Ranomafana National Park, Madagaskar. *Ecological Economics*, 43, 261-275.

Ferraro, P.J., and Pattanayak, S. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, 4:e105.

Fuller, D. O., Jessup, T. C., and Salim, A. 2004. Loss of forest cover in Kalimantan, Indonesia, since the 1997-1998 El Nino. *Conservation Biology*, 18(1), 249-254.

Gan, J. and McCarl, B. 2007. Measuring transnational leakage of forest conservation. *Ecological Economics,* 64(2), 423-432.

Gaveau, D.L., Epting, J., Lyne, O., Linkie, M., Kumara, I., Kanninen, M. and Leader-Williams, N., 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. *Journal of Biogeography*, *36*(11), pp.2165-2175.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), pp.850-853.

Herrera, D., A. Pfaff and J. Robalino. 2015. Spillover Conservation around Protected Areas in the Brazilian Amazon: Frontier Dynamics and Variation in Nearby Deforestation 'Blockage' (Reverse Leakage). Duke University, NC.

Honey-Rosés, J., Baylis, K. and Ramirez, M.I. 2011. Do our conservation programs work? A spatially explicit estimate of avoided forest loss. *Conservation Biology*, 25(5), 1032-

1043.

Imbens, G.M. and Wooldridge, J.M., 2008. *Recent developments in the econometrics of program evaluation* (No. w14251). National Bureau of Economic Research.

International Union for Conservation of Nature (IUCN), U. W. 2010. The world database on protected areas (WDPA). *UNEPWCMC*, Cambridge, UK.

Joppa, L., and Pfaff, A. 2010. Reassessing the forest impacts of protection: the challenge of nonrandom location and a corrective measure. *Annals of the New York Academy of Sciences*, 1185, 135–139.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. 2007. Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97-130.

Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3), 531-542.

Margono, B. A., Turubanova, S., Zhuravleva, I. Potapov, P. Tyukavina, A., Baccini, A., Goetz, S. and Hansen, M. C. 2012. Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environmental Research Letters*, 7(3), 034010. doi:10.1088/1748-9326/7/3/034010.

Mertens, B., and Lambin, E. F. 2000. Land-cover-change trajectories in southern Cameroon. *Annals of the Association of American Geographers*, 90(3), 467-494.

Meyfroidt, P. and Lambin, E. F. 2009. Forest transition in Vietnam and displacement of deforestation abroad. *Proceedings of the National Academy of Science USA*, 106, 16139-16144.

Miettinen, J., Shi, C., H., and Liew, S., C. 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology*, 17, 2261-2270.

Moffitt, R. A. 2001. Policy interventions, low-level equilibria, and social interactions. *Social Dynamics*, 4(45-82), 6-17.

Naidoo, R. and Iwamura, T., 2007. Global-scale mapping of economic benefits from agricultural lands: implications for conservation priorities. *Biological Conservation*, 140, 40-49.

Oliveira, P., Asner, G., Knapp, D. Almeyda, A. Galván-Gildemeister, R., Keene, S. Raybin, and R. Smith, R. 2007. Land-use allocation protects the Peruvian Amazon. *Science*, 317, 1233-1236.

Olsen, N., and Bishop, J. 2009. *The financial costs of REDD: Evidence from Brazil and Indonesia*. Glànd, Switzerland: International Union for Conservation of Nature (IUCN).

Pagiola, S. 2000. Land use change in Indonesia. Background paper prepared for the Environment Department, World Bank, Washington, DC.

Pfaff, A., Robalino, J., Lima, E., Sandoval, C., and Herrera, L. D. 2014. Governance, location and avoided deforestation from protected areas: Greater restrictions can have lower impact, due to differences in location. *World Development*, 55, 7-20.

Pfeifer, M., Burgess, N. D., Swetnam, R. D., Platts, P. J., Willcock, S., and Marchant, R. 2012. Protected areas: mixed success in conserving East Africa's evergreen forests. *PLoS One*, 7(6), e39337.

Robalino, J. A., and Pfaff, A. 2012. Contagious development: Neighbor interactions in deforestation. *Journal of Development Economics*, 97(2), 427-436.

Schwarze, E, J.O. Niles, and J. Olander. 2002. Understanding and managing leakage in forest-based greenhouse-gas-mitigation projects. *Philosophical Transactions of the Royal Society*, 360(1797), 1685-1703.

Shah, P. and Baylis, K. 2015. Evaluating Heterogeneous Conservation Effects of Forest Protection in Indonesia. *PLoS ONE* 10(6): e0124872. doi: 10.1371/journal.pone.0124872.

United Nations. 2016. Forest Crimes. U.N. Office on Drugs and Crime, New York. https://www.unodc.org/indonesia/forest-crime.html

Wear, D. and B. Murray. 2004. Federal timber restrictions, interregional spillovers, and the impact on US softwood markets. *Journal of Environmental Economics and Management,* 47(2), 307-330.

Wich, S., Riswan, J. J., Refisch, J., and Nellemann, C. 2011. *Orangutans and the economics of sustainable forest management in Sumatra. UNEP/GRASP/PanEco*. YEL/ICRAF/GRID Arendal. ISBN 978-82-7701-09.

Wittemyer, G., P. Elsen, W.T., Bean, C.O. Burton, and J. S. Brashares. 2008. Accelerated human population growth at protected area edges. *Science*, 321, 123-126.

Wu, J. 2000. Slippage effects of the conservation reserve program. *American Journal of Agricultural Economics*, 82(4), 979-992.

Wunder, S. 2008. Payments for environmental services and the poor: Concepts and preliminary evidence. *Environment and Development Economics,* 13, 279-97.

**Appendix 1: Linearization of the Basic Model**

To derive the impact of a small increase in $R_E$, we totally differentiate the equations of the general equilibrium model above, and we solve the resulting system of linear differential equations. First, totally differentiating the resource constraints yields:

$$\hat{R}_T\lambda_{RT} + \hat{R}_A\lambda_{RA} + \hat{R}_E\lambda_{RE} = 0 \tag{1}$$

$$\lambda_{LT}\hat{L}_T + \lambda_{LA}\hat{L}_A = \hat{L}_{TA} \tag{2}$$

$$\hat{L}_{TA}\lambda_L + \hat{L}_M\lambda_{LM} = 0 \tag{3}$$

where $\lambda_{Rj}$ is the fraction of $R$ initially in sector $j$ ($j$=$T$,$A$, $E$), so $\lambda_{RT} + \lambda_{RA} + \lambda_{RE} = 1$. Also, $\lambda_{LT}$ and $\lambda_{LA}$ are the fractions of rural labor $L_{TA}$ that are in sectors $T$ and $A$, respectively, so $\lambda_{LT} + \lambda_{LA} = 1$. Finally $\lambda_L$ is the fraction of total labor that is rural, and $\lambda_{LM}$ is the fraction of total labor in sector $M$. Next, we totally differentiate the production functions to find how changes in inputs affect output:

$$\hat{E} = \hat{R}_E \tag{4}$$

$$\hat{T} = \theta_{TR}\hat{R}_T + \theta_{TL}\hat{L}_T \tag{5}$$

$$\hat{A} = \theta_{AR}\hat{R}_A + \theta_{AL}\hat{L}_A \tag{6}$$

$$\hat{M} = \hat{L}_M \tag{7}$$

$$\hat{L} = \theta_{LTA}\hat{L}_{TA} + \theta_{LM}\hat{L}_M \tag{8}$$

where $\theta_{Ti}$ denotes the factor share in sector $T$ for each factor $i$ ($i$=$R$,$L$), so $\theta_{TR} + \theta_{TL} = 1$, and $\theta_{Ai}$ denotes the factor share in sector $A$ for each factor $i$ ($i$=$R$,$L$), so $\theta_{AR} + \theta_{AL} = 1$. Also, $\theta_{LTA}$ and $\theta_{LM}$ are factor shares in the function $L = L(L_{TA}, L_M)$, so $\theta_{LTA} + \theta_{LM} = 1$. Totally differentiating the zero profit conditions yields:

$$\hat{P}_T + \hat{T} = \theta_{TR}(\hat{P}_R + \hat{R}_T) + \theta_{TL}(\hat{w} + \hat{L}_T) \tag{9}$$

$$\hat{P}_A + \hat{A} = \theta_{AR}(\hat{P}_R + \hat{R}_A) + \theta_{AL}(\hat{w} + \hat{L}_A) \tag{10}$$

$$\hat{P}_M + \hat{M} = \hat{w}_M + \hat{L}_M \tag{11}$$

$$\hat{P}_L + \hat{L} = \theta_{LTA}(\hat{w} + \hat{L}_{TA}) + \theta_{LM}(\hat{w}_M + \hat{L}_M) \tag{12}$$

Differentiating the definitions of $\sigma_T$, $\sigma_A$, and $\sigma_L$, and rearranging terms yields:

$$\hat{R}_T - \hat{L}_T = \sigma_T\left(\hat{w} - \hat{P}_R\right) \tag{13}$$

$$\hat{R}_A - \hat{L}_A = \sigma_A\left(\hat{w} - \hat{P}_R\right) \tag{14}$$

$$\hat{L}_{TA} - \hat{L}_M = \sigma_L(\hat{w}_M - \hat{w}) \tag{15}$$

Totally differentiating the budget constraint, $M^C P_M = (R_A + R_T)\, P_R + L_M w_M + L_{TA} w$ yields:

$$\hat{M}^C + \hat{P}_M = \theta_{IR_A}(\hat{R}_A + \hat{P}_R) + \theta_{IR_T}(\hat{R}_T + \hat{P}_R) + \theta_{IL_M}(\hat{L}_M + \hat{w}_M) + \theta_{IL_{TA}}(\hat{L}_{TA} + \hat{w})$$

where $\theta_{IR_A}$, $\theta_{IR_T}$, $\theta_{IL_M}$, and $\theta_{IL_{TA}}$ represent the income shares from $R_A$, $R_T$, $L_M$, and $L_{TA}$, respectively. Thus we have $\theta_{IR_A} + \theta_{IR_T} + \theta_{IL_M} + \theta_{IL_{TA}} = 1$.

Totally differentiating the balance of payments constraint yields:[18]

$$\hat{M}^I + \hat{P}_M = \theta_{BT}(\hat{T} + \hat{P}_T) + \theta_{BA}(\hat{A} + \hat{P}_A) \tag{16}$$

where $\theta_{BT}$ and $\theta_{BA}$ represent the share of export income from sale of goods $T$, and $A$, respectively, and $\theta_{BT} + \theta_{BA} = 1$. We know that the domestic consumption of good $M$ has to equal the sum of domestic production plus imports, $M^C = M + M^I$. Totally differentiating this constraint, we have:

$$\hat{M}^C = \delta_{MD}\hat{M} + \delta_{MI}\hat{M}^I \tag{17}$$

where $\delta_{MD}$ represents the initial share of $M^C$ produced domestically, and $\delta_{MI}$ is the share of $M^C$ imported, and $\delta_{MD} + \delta_{MI} = 1$. Finally, rearranging the definitions of the ROW's demand elasticities implies:

$$\hat{T} = \varepsilon\hat{P}_T \tag{18}$$

$$\hat{A} = \eta\hat{P}_A \tag{19}$$

Since $\hat{R}_E$ is an exogenous policy shock ($\hat{R}_E > 0$), we have 19 equations in 20 unknowns ($\hat{R}_T, \hat{R}_A, \hat{L}_A, \hat{L}_M, \hat{L}_T, \hat{E}, \hat{T}, \hat{M}, \hat{M}^C, \hat{M}^I, \hat{A}, \hat{P}_R, \hat{P}_T, \hat{P}_M, \hat{P}_L, \hat{P}_A, \hat{L}_{TA}, \hat{w}_M, \hat{w}, \hat{L}$). We further assume that labor in $M$ is the numeraire, so $\hat{w}_M = 0$, and we are left with 19 equations for 19 unknowns. Successive substitution is laborious but straightforward to find solutions in the text, including equations (20) through (23).

[18] The combination of the zero profit conditions and the balance of payments constraint together imply the household budget constraint, so we do not use that budget as an independent equation in our system to be solved. If all markets but one are in equilibrium, Walras' Law ensures that the last market must also be in equilibrium.

**Appendix 2: Comparative Statics on Equation (23)**

In section 1.1, equation (23) shows that leakage is $\psi = -\left[\frac{(1-\alpha)}{\alpha} - \frac{\lambda_{RT}N}{\alpha(\lambda_{RT}N+\lambda_{RA}D)}\right]$. Here, we use this equation to prove each theorem.

**Proof of Theorem 1:**

From (23), leakage is negative if and only if $\frac{(1-\alpha)}{\alpha} > \frac{\lambda_{RT}N}{\alpha(\lambda_{RT}N+\lambda_{RA}D)}$. Rearranging, we have:

$$(1-\alpha)(\lambda_{RT}N + \lambda_{RA}D) > \lambda_{RT}N$$

$$\lambda_{RA}D - \alpha(\lambda_{RT}N + \lambda_{RA}D) > 0$$

$$\alpha < \frac{\lambda_{RA}D}{(\lambda_{RT}N + \lambda_{RA}D)}$$

which proves Theorem 1.

Next, we partially differentiate the solution for $\psi$ in (23) with respect to each of the seven parameters: $\alpha$, $\eta$, $\varepsilon$, $\lambda_{RA}$, $\lambda_{RT}$, $\theta_{AL}$, and $\sigma_L$.

**Proof of Theorem 2:**

In order to derive the effect of $\alpha$ on leakage, we take the derivative of equation (23) with respect to $\alpha$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\alpha} = -\frac{\lambda_{RA}D}{\alpha^2(\lambda_{RT}N+\lambda_{RA}D)}$$

Thus, $\frac{\partial\psi}{\partial\alpha} > 0$ if and only if $-\frac{\lambda_{RA}D}{\alpha^2(\lambda_{RT}N+\lambda_{RA}D)} > 0$, or equivalently, $\frac{\lambda_{RA}D}{(\lambda_{RT}N+\lambda_{RA}D)} < 0$.

**Proof of Theorem 3:**

Next, we derive the effect on leakage of the price elasticity of demand of agriculture, $\eta$. We take the derivative of equation (23) with respect to $\eta$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\eta} = \frac{\lambda_{RT}\lambda_{RA}(D(\varepsilon\lambda_{LA}\theta_{AR} - \lambda_{LA}(\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)) - N((\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)\lambda_{LT} + \lambda_{LT}\sigma_T + \lambda_{LA}\sigma_A - \theta_{AR}(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT})))}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2}$$

Here, $\frac{\partial\psi}{\partial\eta} > 0$ if and only if

$$\frac{\lambda_{RT}\lambda_{RA}(D(\varepsilon\lambda_{LA}\theta_{AR} - \lambda_{LA}(\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)) - N((\varepsilon\theta_{TR} - \theta_{TL}\sigma_T)\lambda_{LT} + \lambda_{LT}\sigma_T + \lambda_{LA}\sigma_A - \theta_{AR}(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT})))}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2} > 0.$$

The denominator is positive by definition and $\lambda_{RT}\lambda_{RA} > 0$. So $\frac{\partial\psi}{\partial\eta} > 0$ if and only if

$$D\big(\varepsilon\lambda_{LA}\theta_{AR}-\lambda_{LA}(\varepsilon\theta_{TR}-\theta_{TL}\sigma_T)\big)$$

$$-N\big((\varepsilon\theta_{TR}-\theta_{TL}\sigma_T)\lambda_{LT}+\lambda_{LT}\sigma_T+\lambda_{LA}\sigma_A-\theta_{AR}(\lambda_{LM}\sigma_L+\varepsilon\lambda_{LT})\big)>0.$$

Simplifying the above inequality further, we get: $\frac{\partial\psi}{\partial\eta}>0$ if and only if

$$(\varepsilon(\theta_{TR}-\theta_{AR})-\theta_{TL}\sigma_T)(\lambda_{LA}D+\lambda_{LT}N)+N(\lambda_{LT}\sigma_T+\lambda_{LA}\sigma_A-\lambda_{LM}\sigma_L)<0.$$

**Proof of Theorem 4:**

Next we derive the effect of the price elasticity of demand of timber, $\varepsilon$ on leakage. We take the derivative of equation (23) with respect to $\varepsilon$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\varepsilon}=\frac{\lambda_{RT}\lambda_{RA}\big(D\big(-\theta_{TR}(\eta\lambda_{LA}+\lambda_{LM}\sigma_L)+\lambda_{LT}\sigma_T+\lambda_{LA}\theta_{AR}(\eta+\sigma_A)\big)-N\big(\theta_{TR}\eta\lambda_{LT}-\lambda_{LT}(\eta\theta_{AR}-\theta_{AL}\sigma_A)\big)\big)}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2}$$

The denominator in the RHS is positive, and $\lambda_{RT}\lambda_{RA}>0$, so $\frac{\partial\psi}{\partial\varepsilon}>0$ if and only if:

$$D\big(-\theta_{TR}(\eta\lambda_{LA}+\lambda_{LM}\sigma_L)+\lambda_{LT}\sigma_T+\lambda_{LA}\theta_{AR}(\eta+\sigma_A)\big)-N\big(\theta_{TR}\eta\lambda_{LT}-\lambda_{LT}(\eta\theta_{AR}-\theta_{AL}\sigma_A)\big)>0.$$

**Proof of Theorem 5:**

In order to derive the effect of the initial land allocated for agriculture, $\lambda_{RA}$ on leakage, we take the derivative of equation (23) with respect to $\lambda_{RA}$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\lambda_{RA}}=-\frac{ND(\lambda_{RT}+\lambda_{RA})}{\alpha\lambda_{RE}(\lambda_{RT}N+\lambda_{RA}D)^2}$$

The denominator in the RHS of this equation is positive and $\lambda_{RT}+\lambda_{RA}>0$ so $\frac{\partial\psi}{\partial\lambda_{RA}}>0$ if and only if:

$$ND<0.$$

**Proof of Theorem 6:**

In order to derive the effect on leakage of the initial land allocated for timber, $\lambda_{RT}$, we take the derivative of equation (23) with respect to $\lambda_{RT}$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\lambda_{RT}}=\frac{ND(\lambda_{RT}+\lambda_{RA})}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2}.$$

The denominator in the RHS of this equation is positive and $(\lambda_{RT}+\lambda_{RA})>0$ so $\frac{\partial\psi}{\partial\lambda_{RT}}>0$ if and only if:

$$ND>0.$$

**Proof of Theorem 7:**

In order to derive the effect on leakage of the factor share of land in agriculture, $\theta_{AL}$, we take the derivative of equation (23) with respect to $\theta_{AL}$, and rearrange, to get:

$$\frac{\partial\psi}{\partial\theta_{AL}} = \frac{-\lambda_{RT}\lambda_{RA}(\eta+\sigma_A)\left(D\varepsilon\lambda_{LA}+N(\lambda_{LM}\sigma_L+\varepsilon\lambda_{LT})\right)}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2}$$

The denominator in the RHS is positive, and $-\lambda_{RT}\lambda_{RA} < 0$, so $\frac{\partial\psi}{\partial\theta_{AL}} < 0$ if and only if:

$$(\eta+\sigma_A)(D\varepsilon\lambda_{LA} + N(\lambda_{LM}\sigma_L + \varepsilon\lambda_{LT})\ ) > 0.$$

**Proof Theorem 8:** To derive the effect on leakage of labor mobility between the rural sector and urban sector, we take derivative of leakage with respect to $\sigma_L$:

$$\frac{\partial\psi}{\partial\sigma_L} = \frac{\lambda_{RT}\lambda_{RA}\lambda_{LM}(N(\eta\theta_{AR}-\theta_{AL}\sigma_A)-D(\varepsilon\theta_{TR}-\theta_{TL}\sigma_T))}{\alpha(\lambda_{RT}N+\lambda_{RA}D)^2}.$$

The denominator in the RHS is positive, and $\lambda_{RT}\lambda_{RA}\lambda_{LM} > 0$, so $\frac{\partial\psi}{\partial\sigma_L} > 0$ if and only if:

$$N(\eta\theta_{AR} - \theta_{AL}\sigma_A) - D(\varepsilon\theta_{TR} - \theta_{TL}\sigma_T) > 0.$$

**Appendix 3: Analysis based on 35km Distance to Delineate "Nearby" Areas**

The text measures leakage in districts near new parks, whereas other previous studies measure leakage within a specified distance (Andam *et al.* 2008; Pfeifer *et al.* 2012). In this appendix, we also delineate a 35 km zone around newly established protected areas within which to estimate the effect of the key economic variables on the change in forest cover (relative to controls further away). Table A3-1 presents the summary statistics for characteristics of parcels that are within and outside this 35 km zone.

**Table A3-1: Summary Statistics for Parcels inside and outside 35km Treatment Zone**

| Variable | 35km Treatment Zone Mean | 35km Treatment Zone Std. Dev. | Parcels outside those Zones Mean | Parcels outside those Zones Std. Dev. |
|---|---|---|---|---|
| Neighbors' Average Forest Loss (ha) | 122.96 | 151.49 | 79.40 | 128.06 |
| Forest Loss (ha) | 124.92 | 186.28 | 81.41 | 154.02 |
| Neighbors' Average Distance to City (km) | 68.88 | 64.39 | 125.52 | 98.77 |
| Distance to City (km) | 71.42 | 70.45 | 128.39 | 99.86 |
| Neighbors' Avg Distance to Road (km) | 21.38 | 29.36 | 58.88 | 82.37 |
| Distance to Road (km) | 21.75 | 29.73 | 60.69 | 85.38 |
| Neighbors' Average Elevation (m) | 289.69 | 359.52 | 302.40 | 445.68 |
| Elevation (m) | 287.07 | 369.18 | 304.91 | 479.90 |
| Neighbors' Average Slope (deg) | 6.03 | 5.31 | 6.78 | 6.33 |
| Slope (deg) | 6.01 | 5.76 | 6.88 | 6.79 |
| No. of Observations | 6856 | | 164043 | |

Notes for this table correspond to notes on Table 4 of the text.

We identify counterfactual parcels that are outside of this 35 km zone around the new protected areas but that are similar to the parcels within this zone. Table A3-2 presents the summary statistics for the key economic variables. Table A3-3 shows normalized differences for the covariates used in the matching process for this matched dataset. Table A3-4 presents the summary statistics for other land characteristics that are not used in the matching process.

**Table A3-2: Summary Statistics for Key Economic Variables in the Matched Dataset (for 35km Treatment Zone)**

| Variable | Treatment Parcels Mean | Treatment Parcels Std. Dev. | Control Parcels Mean | Control Parcels Std. Dev. |
|---|---|---|---|---|
| Share of new park taken from $A$ ($\alpha$) | 0.01 | 0.08 | 0.00 | 0.00 |
| Elasticity of demand for $A$ ($\eta$) | 0.49 | 0.14 | 0.50 | 0.13 |
| Elasticity of demand for $T$ ($\varepsilon$) | 0.58 | 0.42 | 0.54 | 0.37 |
| Initial fraction of land in $A$ ($\lambda_{RA}$) | 0.15 | 0.28 | 0.20 | 0.35 |
| Initial fraction of land in $T$ ($\lambda_{RT}$) | 0.42 | 0.42 | 0.34 | 0.42 |
| Ag employment/Ag GRDP ($\theta_{AL}$) | 1.69 | 1.01 | 1.85 | 1.20 |
| Log of wage rate ($\sigma_L$) | 13.11 | 0.28 | 13.12 | 0.27 |

Notes for this table correspond to notes on Table 5 of the text.

**Table A3-3: Covariate Balance for Variables Used in the Matching Process (for 35km Treatment Zone)**

| | Treatment Parcels | | Control Parcels | | |
|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Norm. Diff** |
| Neighbors' Average Forest Loss (ha) | 124.84 | 152.33 | 122.09 | 136.40 | 0.01 |
| Forest Loss (ha) | 126.77 | 187.44 | 122.93 | 169.15 | 0.02 |
| Neighbors' Avg Distance to City (km) | 69.51 | 64.82 | 68.14 | 62.72 | 0.02 |
| Distance to City (km) | 72.10 | 70.94 | 70.53 | 68.44 | 0.02 |
| Neighbors' Avg Distance to Road (km) | 21.67 | 29.60 | 19.49 | 25.87 | 0.06 |
| Distance to Road (km) | 22.06 | 29.97 | 19.77 | 26.14 | 0.06 |
| Neighbors' Average Elevation (m) | 272.09 | 341.09 | 246.27 | 324.63 | 0.05 |
| Elevation (m) | 269.89 | 352.21 | 242.95 | 332.85 | 0.06 |
| Neighbors' Average Slope (deg) | 5.93 | 5.30 | 5.70 | 5.09 | 0.03 |
| Slope (deg) | 5.92 | 5.74 | 5.68 | 5.49 | 0.03 |

Notes for this table correspond to notes on Table 7a of the text.

**Table A3-4: Covariate Balance for Variables Not Used in the Matching Process (for 35km Treatment Zone)**

| | Treatment Parcels | | Control Parcels | | |
|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Norm. Diff** |
| Distance to River (m) | 3144.75 | 3653.45 | 2965.68 | 3158.53 | 0.04 |
| Neighbors' Avg Distance to River (m) | 3091.71 | 3277.22 | 2949.59 | 2850.01 | 0.03 |
| Accessibility (minutes) | 579.79 | 592.98 | 471.56 | 518.55 | 0.14 |

Notes for this table correspond to notes on Table 7b of the text.

**Table A3-5: Key Coefficients from Individual Regressions of Forest Loss on Covariates and One Key Variable at a Time**

| Economic Parameter from Theory Model | Expected Sign from Theory Model | Proxy Economic Variable | Coefficient Estimate from Individual OLS Regression (clustered standard errors) | Marginal Effect from Individual IV Spatial Lag Regressions (robust standard errors) |
|---|---|---|---|---|
| Share of park taken from *A* ($\alpha$) | + | Treatment × Share of park taken from *A* | 13.14<br>(29.88) | 406.62<br>(271.57) |
| Elasticity of demand for *A* ($\eta$) | - | Treatment × Elasticity of *A* | -118.52***<br>(19.89) | -362.89**<br>(173.61) |
| Elasiticity of demand for *T* ($\varepsilon$) | + | Treatment × Elasticity of *T* | 13.96**<br>(7.06) | 68.21<br>(51.70) |
| Initial fraction of land in *A* ($\lambda_{RA}$) | - | Treatment × Fraction of land in *A* | -9.36<br>(11.60) | 175.20<br>(113.07) |
| Initial fraction of land in *T* ($\lambda_{RT}$) | + | Treatment × Fraction of land in *T* | 57.80***<br>(4.88) | 139.85***<br>(52.69) |
| Factor share of labor in agriculture ($\theta_{AL}$) | - | Treatment × Ratio of ag. workers /ag. output | 6.63***<br>(2.57) | 18.45<br>(22.18) |
| Labor mobility between rural (*A* and *T*) vs urban (*M*) sector ($\sigma_L$) | - | Treatment × Log of wage rate | -3.35<br>(9.86) | -42.98<br>(70.27) |

*** p<0.01, ** p<0.05, * p<0.1

**Table A3-6: Regressions of Forest Loss on All Variables Together (t-statistic in parentheses)**

| | **OLS** | **Instrumental Variable** |
|---|---|---|
| Spatial lag in forest cover loss 2000-2012 | | |
| Treatment dummy | -105.56 | 559.99 |
| | (158.46) | (1101.44) |
| Distance to city (km) | 0.72*** | 0.07 |
| | (0.07) | (0.52) |
| Distance to road (km) | 0.71*** | 2.13* |
| | (0.15) | (1.12) |
| Elevation (m) | -0.04*** | -0.06** |
| | (0.00) | (0.03) |
| Slope (deg) | -7.06*** | -11.58*** |
| | (0.34) | (2.82) |
| Elasticity of demand for *A* | 56.18*** | 286.11** |
| | (16.42) | (138.95) |
| Elasticity of demand for *T* | -22.94*** | -65.15* |
| | (6.42) | (39.45) |
| Fraction of total land initially in *A* | 20.68** | -78.23 |
| | (9.22) | (74.45) |
| Ratio of employment in *A* to district-level GRDP from *A* | -19.51*** | -11.46 |
| | (2.12) | (12.56) |
| Log of wage rate | -12.73 | -15.46 |
| | (11.54) | (59.90) |
| Treatment × $\alpha$ (Share of park taken from *A*) | 24.25 | 369.81 |
| | (29.04) | (250.85) |
| Treatment × $\eta$ (Elasticity of demand for *A*) | -118.89*** | -386.78** |
| | (20.41) | (176.72) |
| Treatment × $\varepsilon$ (Elasticity of demand for *T*) | 18.77** | 85.53 |
| | (7.54) | (54.02) |
| Treatment × $\lambda_{RA}$ (Fraction of land in *A*) | 0.09 | 232.51* |
| | (13.10) | (126.34) |
| Treatment × $\theta_{LA}$ (Ratio of employment in *A* to district-level GRDP of *A*) | 6.17** | 37.69 |
| | (3.10) | (24.83) |
| Treatment × $\sigma_L$ (Log of wage rate) | 11.09 | -37.95 |
| | (11.76) | (80.91) |
| **Regional FE** | yes | yes |
| **Observations** | 40,276 | 40,276 |
| **R-squared** | 0.2394 | 0.7254 |

*** p<0.01, ** p<0.05, * p<0.1